Towards tutoring systems that detect students' motivation: an investigation

Angel de Vicente

Doctor of Philosophy

Institute for Communicating and Collaborative Systems
School of Informatics
University of Edinburgh

2003

Abstract

The use of Artificial Intelligence techniques in the development of educational software brought the hope of developing systems that would become personalised to each student and thus be of more benefit to him or her. But despite their added complexity, these “Intelligent” systems (ITSs, ILEs, ICALLs, etc.) do not always succeed in engaging the student. While a lot of effort has been spent investigating how to accommodate an instructional interaction to the student’s knowledge, almost no work has been done in trying to accommodate the instruction to the student’s motivational state. This is surprising, given the immense impact that a student’s motivation has in his or her learning.

The little previous work dealing explicitly with motivation in tutoring systems has focused mainly on the strategies that an Intelligent Tutoring System (ITS) could use to motivate the student. In this dissertation we focus on the prior (but we believe, fundamental), task of detecting the student’s motivational state, on which the mentioned strategies could be used.

We argue that the available theories of motivation in education are not specific enough and are of limited usefulness in order to implement a motivation detection component in an ITS. Thus, we argue for the need of empirical studies that can help us elicit formalised motivation diagnosis knowledge. To this effect, we discuss a number of empirical studies we performed in order to inform the design of an ITS simulation that detects the motivational state of a student.

The main aspects of the motivation diagnosis architecture presented in this dissertation are a motivation self-report component and a motivation diagnosis component based on human teachers’ motivation diagnosis knowledge, elicited via one of the mentioned empirical studies. This architecture was implemented as an ITS simulation in order to help us evaluate these motivation diagnosis techniques.

The evaluation showed that, although not perfect, the motivation diagnosis techniques introduced in this dissertation seem to offer a reasonable level of accuracy in detecting a student’s motivational state, and although the approach presented is not the only possible one and many aspects of this work can still be improved, we believe that it offers a promising step towards tutoring systems that care!

Acknowledgements

And I have by me, for my comfort, two strange white flowers [...] to witness that even when mind and strength had gone, gratitude and a mutual tenderness still lived on in the heart of man.
Time Machine — Herbert George Wells

In the *working arena* I would like, first of all, to thank the Spanish Government for generously providing my bread and butter[1] during most of the length of the project. Locally, I am indebted to the School of Artificial Intelligence, and more specifically to the Artificial Intelligence and Education (AI-ED) group, for providing a rich and stimulating working environment, and for very useful and encouraging ideas for this work.

Much of the knowledge needed to create an aesthetically pleasant document I owe to Rafael Morales, the local LaTeX expert. I would also like to thank Madoka Aiki for checking that the Japanese used for the ITS prototype developed was correct. Most of all, I would like to thank my supervisor, Helen Pain, for her support throughout and for her understanding and patience during my lazy moments (which were perhaps too many).

In the *personal arena* I also have to thank the Artificial Intelligence and Education (AI-ED) group for numerous social activities which made the Scottish weather more wearable. Although far away, my family was always there supporting me, and in return I showed them part of this beautiful country.

I owe immense gratitude to many friends who have suffered some of my laments and complaints during this time, and who have exchanged these for encouragement, support, fun, friendship and love in huge quantities. The deepest thanks for bearing with me go to Yolanda, who helped me to keep going through rough times.

Many more people should be included in these acknowledgements, and I apologise to all of them for not having the space to include them all. Thanks to all of you who have made my time in Edinburgh truly unforgettable.

[1]Thanks to grant PG 94 30660835 from the "Programa de Becas de Formación de Profesorado y Personal Investigador en el Extranjero" of the Ministerio de Educación y Cultura.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified. Parts of the work described in this thesis have been previously published and are reproduced in appendix E with permission.

(Angel de Vicente)

To my father — who could not see this work finished.

Table of Contents

1 Introduction **1**

1.1 Intelligent Tutoring Systems 2
1.2 How to create an Affective Tutor? 4
1.3 Research goals 6
1.4 Structure of the dissertation 7

2 Background and Related Work **11**

2.1 Introduction 11
2.2 The terminological forest 13
2.2.1 Emotions 14
2.2.2 Motivation 15
2.3 Affective Computing 17
2.3.1 Emotion detection 18
2.3.2 Emotion synthesis (or simulation) 20
2.3.3 Generation of artificial emotions 22
2.3.4 Conclusion 23
2.4 Motivation in Education 23
2.4.1 A theory of motivation in Education 24
2.4.2 How to influence student's motivation? 27
2.5 Affective Tutors 29
2.5.1 Implicitly affective tutors 30
2.5.2 Relevant issues for affective tutors 32
2.5.3 Affective tutors explicitly 37

2.6 Conclusions 40

3 Outlining the design of an Affective Tutor **43**

3.1 Motivation modelling 43

3.2 *Which* motivational factors to model? 44

3.3 *How* to diagnose those motivational factors? 48

3.4 Putting it all together 50

4 Self-report study **51**

4.1 Goal of study 51

4.2 Materials 51

4.3 Methodology 55

4.4 Results 57

4.4.1 On acceptance of the self-report method 57

4.4.2 How often and when were the sliders updated? 58

4.4.3 Which values did the sliders take? 64

4.5 Discussion 71

4.5.1 Suggested changes to the self-report interface 73

5 Motivation Diagnosis study **75**

5.1 Introduction and goal of the study 75

5.2 Materials 77

5.2.1 A_MOODS description 77

5.3 Methodology 80

5.4 Results 81

5.4.1 Two excerpts in detail 83

5.4.2 Elicited motivation diagnosis knowledge 87

5.5 Discussion 104

6 Motivation Diagnosis Validity Study **107**

6.1 Introduction and goal of the study 107

6.2 Materials 108

6.3 Participants 110

6.4 Methodology 110
6.5 Results 111
6.5.1 Was it easy to select an option? 112
6.5.2 Which rules to accept? 113
6.5.3 Comments by participants 115
6.6 Discussion 120

7 MOODS. A prototype Affective Tutor Simulation 123
7.1 Introduction 123
7.2 MOODS Structure Overview 124
7.3 Student models 127
7.3.1 Traits model 127
7.3.2 Reported, Inferred and Combined motivation models 127
7.4 Knowledge-based rules 128
7.4.1 Incorporating the instructional planning rules from del Soldato (1994) 128
7.4.2 Motivational planning rules 133
7.4.3 Affective Dialogue rules 136
7.4.4 Motivation diagnosis rules 141
7.4.5 Combination rules 142
7.5 Interaction history 143
7.6 Discussion 143

8 Example interaction 145
8.1 Introduction 145
8.2 MOODS Interface 145
8.2.1 Variables frame 147
8.3 Instructional Simulation example 148
8.3.1 Motivation diagnosis inferences 151
8.4 Discussion 152

9 Evaluation 155
9.1 Introduction and goal of evaluation 155

9.2 Materials 156
9.2.1 Simulations 158
9.3 Participants 162
9.4 Methodology 163
9.5 Results 164
9.5.1 Was it easy for participants to select a value? 166
9.5.2 Did the participants agree on their inferences of the motivational model values? 168
9.5.3 How does MOODS compare with the inferences made by the participants? 169
9.5.4 How important were the motivational factors considered? . . . 177
9.6 Discussion 180

10 Conclusions and Further Work **183**
10.1 Summary of research and findings 183
10.2 Some criticisms 188
10.3 Further work 192
10.4 Conclusions 194

Bibliography **197**

A Self-report study materials **209**
A.1 Prototype ITS 210
A.1.1 Instructional domain 210
A.1.2 Usage description 213
A.2 Materials given to students prior to interaction. 216
A.3 Post-questionnaire given to participants 220

B Motivation Diagnosis study materials **223**
B.1 Instructions for participants. 224
B.2 Complete transcription of one interview 232
B.3 Preliminary rules for motivation diagnosis study 254

C Validity study materials **271**
C.1 Questionnaire instructions 272
C.2 Complete Questionnaire 279
C.3 Validity study detailed results 288

D Evaluation materials **295**
D.1 Questionnaire instructions 296
D.2 Complete questionnaire 302
D.2.1 Simulation 1. 302
D.2.2 Simulation 2. 303
D.3 Motivational model question results 306
D.4 Importance question results 310

E Published papers **315**

List of Figures

2.1 A model of motivation, performance, and instructional influence (reproduced from Keller, 1983, p. 392). 26

3.1 Motivation model 47

4.1 Main interface of prototype ITS system 52
4.2 Relation between self-report acceptance and the number of slider updates 60
4.3 Distribution of slider update totals through duration of lessons. 61
4.4 Distribution of individual sliders updates through duration of lessons. 62
4.5 Total number of updates per value of slider scale 65
4.6 Sliders' average values for the six instructional paths 68
4.7 The effect of instructional paths on participants' motivational state . . 70

5.1 A_MOODS interface. 77
5.2 Sample message to inform the participant of time spent in theory lesson. 79
5.3 Sample message to remind the participant to update the model. 79
5.4 Inference rule from Excerpt 1. 84
5.5 Inference rule from Excerpt 2. 86
5.6 Occurrences for each input category. 91
5.7 Occurrences for each output category. 92

6.1 Validity Study Sample question. 108
6.2 Teaching experience of participants in Validity Study. 111
6.3 Distribution of *Don't know* replies (for the 31 participants that completed the 30 questions). 113
6.4 Acceptance distribution of Motivation Diagnosis rules. 115

6.5 Number of comments per participant (for the 31 participants that completed the 30 questions). 117

7.1 MOODS Structure Overview 125
7.2 Relationship between preconditions and actions of the rules used to derive *Sol1*. 132

8.1 MOODS Interface 146
8.2 Sample interaction with MOODS 150

9.1 Evaluation Study sample question. 157
9.2 Teaching experience of participants in Evaluation Study. 164
9.3 Distribution of *Don't know* replies for the two types of questions. . . . 166
9.4 Distribution of *Don't know* replies for motivational model question by factors. 167
9.5 Distribution of standard deviations for questions in Simulations 1 and 2. 170
9.6 Average values of motivational factors for all instructional situations . 171
9.7 Values of motivational factors inferred by MOODS for all instructional situations 173
9.8 Distribution of the differences between the values inferred by MOODS and the average values by the study participants. 174
9.9 Distribution of the differences between the values inferred by MOODS and the average values of the study participants (in standard deviations from the mean). 176
9.10 Distribution of standard deviations for *Importance* question in both simulations. 178
9.11 Average values of *Importance* question for all motivational factors and instructional situations. 179

A.1 Main interface 212
A.2 Numbers game 213
A.3 Trait characteristics questionnaire 214
A.4 Main user interface 217

B.1 Interface for study. 231
B.2 Example of lesson. 231
B.3 Sample message to inform of time spent in theory lesson. 231
B.4 Sample message to remind of updating model. 231

List of Tables

2.1 Factors influencing student's motivation (Malone and Lepper, 1987). . 28
2.2 Some rules for confidence modelling, from (del Soldato, 1994, Table 4.7, p. 39) 38
2.3 Motivational Planner—Teacher's actions when Student succeeds in solving a problem. Adapted from (del Soldato, 1994, Table 3.2, p. 28) . . 39
2.4 First two rules of the negotiation planner, from (del Soldato, 1994, Table 4.12, p. 46) 41

3.1 Definitions of motivation model variables (traits and states) 46

4.1 Instructional paths 53
4.2 Description of instructional units 54
4.3 Slider usage 58
4.4 Main times of motivation sliders updates 63

5.1 Basic statistics 82
5.2 Graphical conventions for table 5.3. 87
5.3 Examples of motivation diagnosis rules 88
5.4 Increase Satisfaction diagnosis rules 94
5.5 Decrease Satisfaction diagnosis rules 95
5.6 Increase Confidence diagnosis rules 96
5.7 Decrease Confidence diagnosis rules 97
5.8 Increase Effort diagnosis rules 98
5.9 Decrease Effort diagnosis rules 98
5.10 Increase Cognitive Interest diagnosis rules 99

5.11 Decrease Cognitive Interest diagnosis rules 99
5.12 Increase Sensory Interest diagnosis rules 100
5.13 Decrease Sensory Interest diagnosis rules 100

6.1 Rule IE6 (Increase Effort) (*Reproduced from table 5.8).* 109
6.2 Sample motivation diagnosis rules results. 112
6.3 Sample chi-square results. 114
6.4 Rules to accept. (p<0.01) 116
6.5 Sample comments. 118

7.1 Instructional Planning rules adapted from del Soldato (1994) 130
7.2 Rules from del Soldato (1994) used to derive rule *Sol1*. 131
7.3 Motivational Planning rules. 135
7.4 Affective Dialogue (AFDI) rules. 139
7.5 Surface realization of AFDI rules and the effect of a student's reply on his motivational state. 140
7.6 Increase Effort diagnosis rules (reproduced from table 5.8). 142

8.1 Number of preconditions met for some of the Satisfaction detection rules. 152

9.1 Occurrence of instructional variables in motivational diagnosis rules in MOODS. 160
9.2 Example of results for motivation model question. 165
9.3 Example of results for importance question. 165
9.4 Example of standard deviations as a measure of agreement. 169
9.5 Standard deviations of motivational factors for all instructional situations 172
9.6 Chi-square test 175

A.1 Description of instructional units 211
A.2 DONE feedback 215
A.3 GIVE UP feedback 215

B.1 Graphical conventions for tables B.2 to B.11 254
B.2 Motivation diagnosis rules of participant 1 255

B.3 Motivation diagnosis rules of participant 2 256
B.4 Motivation diagnosis rules of participant 3 258
B.5 Motivation diagnosis rules of participant 4 259
B.6 Motivation diagnosis rules of participant 5 261
B.7 Motivation diagnosis rules of participant 6 261
B.8 Motivation diagnosis rules of participant 7 263
B.9 Motivation diagnosis rules of participant 8 264
B.10 Motivation diagnosis rules of participant 9 266
B.11 Motivation diagnosis rules of participant 10 267

C.1 Results of motivation diagnosis questionnaire. 288
C.2 Chi-square results for all the rules. 290
C.3 Rules on which null hypothesis cannot be rejected. 293
C.4 Null hypothesis can be rejected, but rules cannot be accepted. 294

D.1 Results of motivational model question, Simulation 1. 306
D.2 Results of motivational model question, Simulation 2. 308
D.3 Results of importance question, Simulation 1. 310
D.4 Results of importance question, Simulation 2. 312

Chapter 1

Introduction

Teachers open the door, but you must enter by yourself.
Chinese Proverb

Learning is a crucial aspect of our life: in varying degrees, we all spend an enormous part of it acquiring new knowledge. How successful we are at this life-long task depends on many factors, a very important one being our teachers. We cannot reasonably expect that a teacher will always turn a uninterested pupil into a successful learner. But we can identify a 'good' teacher by how wide (using the same metaphor as in the Chinese proverb above) she[1] opens the door for her students.

Although knowledge of the subject being taught is a prerequisite, a measure of the 'goodness' of a teacher cannot be based solely on her mastery of the material being taught. The necessary qualities for a successful teacher are many but we can categorise them in abstract terms as being able to recognise the particular abilities, shortcomings, preferences, etc. of her students; and to be able to exploit those characteristics in order to successfully direct her students towards the desired goal. This is, obviously, very general, as the methods for recognising and exploiting the student's characteristics could be many, and student's characteristics themselves are of many different types. Thus, abilities do not only refer to intellectual capabilities, but also, for example, to

[1]The use of pronouns of specific gender when referring to a generic person is usually a controversial issue. In this dissertation sexist language will play a useful role, as it will help to identify the two main generic characters in it: the teacher and the student. In a somehow arbitrary fashion, we have decided to consider the gender of the teacher as female, and that of the student as male.

social skills and tenacity.

Even the teacher's desired goal itself is not an obvious choice. In many instructional programmes, student success (i.e. high performance) seems to be the ultimate goal, but this emphasis on high performance has its own risks. As Clifford puts it: "There is reason to speculate that success has become an end in itself, has been giving priority over learning, and may actually be a deterrent to learning." (Clifford, 1990, p. 62).

Given this huge range of issues that a successful teacher has to take into account and given that we still know very little about how a teacher should behave to become a 'good' teacher, it is no big surprise that the success of computerised instructional systems has been very limited. This is not to say that all existing systems are not useful, but rather that the most useful ones developed to date have to be treated as mere tools to be incorporated in a primarily human-to-human instruction.

1.1 Intelligent Tutoring Systems

Throughout the years, many Computer Assisted Instruction (CAI) systems have been developed. From these, only a few attempt to make explicit use of knowledge about human teaching and learning in order to create an instructional situation that adapts to the characteristics of its user. These systems are generically known as Intelligent Tutoring System (ITS)[2], and are differentiated from CAI systems by showing their "intelligence" through the implementation of at least one of these three modules: the domain module, the tutorial module and the student module.

The domain module contains the knowledge about the domain; the student module contains information about the student (for instance: how much he knows, how many times he has used the system, etc.); and the tutorial module contains the knowledge about tutoring (for instance: in which order to present the lessons, what type of feedback to give, etc.). Many examples of this architecture can be found in classic

[2]Though there is a growing number of related terms that refer to systems whose emphasis is on a particular method of instruction or on a particular subject domain (e.g. Intelligent Learning Environments (ILE), Computer-Supported Collaborative Learning (CSCL), Intelligent Computer Assisted Language Learning (ICALL)).

introductory books to the field (e.g. Polson and Richardson, 1988; Wenger, 1987).

Some interesting research has been done in understanding how and what should be implemented in each of these modules, and how these could be augmented in order to increase their efficiency. Perhaps one of the most interesting suggestions in this direction refers to the importance of being able to incorporate into a tutoring system a model of how affective characteristics influence teaching and learning. It is surprising how little research in the field of Artificial Intelligence and Education (AI-ED) has focused in this issue, despite its importance in 'traditional' Education.

How emotional life affects students' motivation to learn and how to influence the latter is a basic concern in classroom practice. Taking this into account for the development of ITSs could bring many benefits. For instance, an ITS could focus on different aspects of the instruction depending on personality characteristics of the student; it could treat mistakes as less important if the student is going through a particularly bad time; or it could attempt to put the student in a particular mood in which he could be more receptive to the material being taught.

But then, given its importance and benefits, why has so little research in AI-ED focused on affective issues? One of the main reasons is probably its difficulty. The worst scenario for a project can be summarised in three points:

1. We do not understand the theoretical aspects of the project.

2. Even if we understood them, the implementation would be too difficult.

3. Even if the implementation was done, it would be difficult to know if we succeeded.

Somehow, the issue of developing ITSs that take into account the motivational state of its student (Affective Tutors, for brevity) shares a similar scenario:

1. Theories of motivation in Education are too vague and/or contradictory, and their implications for actual teaching practice are not always straightforward.

2. The components required for creating an Affective Tutor cannot be developed in isolation. They require the development of an actual tutoring system, but one

whose functioning has been radically changed by introducing components that deal with the motivational aspects of instruction.

3. How can we know when we have successfully created an Affective Tutor? Measuring exclusively performance or enjoyment would certainly be inadequate. Ideally, we would like a composite measure of satisfaction and performance, but how should we weight each of them? And how could we compare this ITS with previously developed systems?

Despite this somewhat bleak picture, it is on the development of Affective Tutors that we focus in this dissertation. As Sloman mentioned when discussing the related issue of developing machines which could understand and communicate affect in language:

> I think it is clear that giving machines all, or even most, of these abilities will be well beyond the state of the art for many years to come. But it is important to keep trying, both as one of many ways to increase our self understanding and because there may be worthwhile practical results. At the very least, studying the problems may give us clues as to how to remedy some of the many deficiencies in communication between people.
>
> (Sloman, 1992, p. 258)

1.2 How to create an Affective Tutor?

How can we develop an Affective Tutor? Which changes and additions to the traditional ITS architecture are needed? Looking first at the three basic modules of an ITS mentioned above, we see that such changes should affect all of them:

- The tutorial module contains the pedagogical knowledge, and traditional tutorial modules should be augmented with knowledge about the motivational influence that specific instructional actions will have on the student. In order to bring about the motivational influences required by its educational goals, this module should have information about the current affective state of the student, and have at its disposal a number of curriculum options from which to choose the one that has the desired motivational influence. This would also require changes in the domain and the student modules.

- Changes to the traditional student module are crucial for the development of an Affective Tutor. The Affective Tutor should keep a model of the student that represents not only his cognitive state, but also his motivational state, dynamically diagnosed during the interaction with the system. This is, in itself, a very difficult problem which should take into consideration many aspects of the student's personality, the curriculum difficulty, performance measures, etc.

- The domain module should also be changed to allow for an extension into the affective realm of what Ohlsson called 'The Principle of Generative Interfaces', i.e. being "able to generate different presentations of each subject-matter unit as needed, at each moment in time choosing the form which is most beneficial for the learner at that moment" (Ohlsson, 1987, p. 217). Thus, should we, at a given point in the interaction, present the next instructional unit as a drill?, as a game?, as a directed unit in which the student does not actively participate? The choice between all these possibilities should be taken by the tutorial module, but we should provide a flexible domain module that allows our system to make use of different presentation alternatives.

As we can see, the development of an Affective Tutor touches almost every aspect of an ITS. But, to date, there has been little research done addressing these issues. The most relevant work is that of del Soldato (1994). She developed a system that uses a combination of traditional domain and motivation-based techniques to plan the instruction. The domain-based planner suggests actions aiming to advance across the domain, while the motivational planner suggests tactics to increase or maintain the student's motivation to work. A negotiation planner attempts to keep an adequate balance between the tactics (sometimes contradictory) suggested by the domain-based and the motivational planner. But one of the issues not greatly explored in her work was that of the detection of the motivational state of a student, on which we focus our work.

1.3 Research goals

As mentioned above, the focus of this dissertation is on the detection of students' motivation, and the main three research goals are:

1. To explore issues of student motivation diagnosis: what methods can we use? Which one seems to be the most adequate one? How do we implement them?

   This goal is addressed in this dissertation by reviewing a large body of research work on issues of detection of emotions and motivation in chapter 2. We also explore in detail the use of self-report in chapter 4 and the possibility of eliciting knowledge-based rules for motivation detection based on teachers' expert knowledge in chapter 5.

2. To explore ways of formalizing knowledge about motivation detection based on empirical studies of instructional analysis. Given that theories of motivation tend to be too general for practical use, can we find ways in which empirical studies can inform the design of Affective Tutors?

   This goal is addressed in this dissertation by presenting three empirical studies we devised and carried out in order to:

   (a) elicit teachers' expert knowledge (chapter 5),

   (b) validate the previously elicited knowledge (chapter 6),

   (c) evaluate the validated knowledge for its accuracy for motivation detection (chapter 9).

3. To explore the issue of Affective Tutors evaluation. How can we perform a systematic evaluation of a tutor in the motivational realm? What type of considerations should we take into account to consider one Affective Tutor more "successful" than another?

   The issue of evaluating an Affective Tutor is addressed in chapter 9, in which we present an evaluation of an instructional interaction simulation (introduced in chapter 7), with which we were able to evaluate the knowledge elicited from the previously mentioned empirical studies against the performance of teachers.

1.4 Structure of the dissertation

The dissertation is structured as follows:

Chapter 1. Where we introduce the problem to be addressed.

Chapter 2. In this chapter we review the work relating to emotions and motivation in general, and more specifically to motivation in education and to motivation in tutoring systems. We see that there has been a variety of different approaches to issues like the detection by a computer of a person's emotions but we also see that research dealing explicitly with motivation in ITSs is very limited.

Chapter 3. Here we present an outline of the design of a tutoring system that could detect the motivational state of a student, based on the theories of motivation in education and the work reviewed in chapter 2. We present a model of motivation based on a number of theories of motivation in education, which will be used to represent the motivational state of a student, and we introduce the methods for motivation detection that we will explore in this dissertation.

Chapter 4. Guided by the design outlined in chapter 3, in this chapter we explore the appropriateness of using a motivation self-report component in a tutoring system. We discuss an empirical study that we performed in order to:

1. investigate whether self-report could really be a viable choice as a motivation diagnosis technique;
2. and if this was the case, to find out which approach to self-report should be the most appropriate to present to a student.

Chapter 5. In this chapter we explore how to elicit knowledge about motivation detection from human teachers by describing an empirical study in which teachers watched the interaction of students with a tutoring system and were asked to infer the likely changes in the motivational state of the students and to formalize their reasoning in order to create a set of motivation diagnosis rules that could be easily implemented in a tutoring system. As a result of this study, we collected a large number of rules regarding the detection of a student's motivational state, which are validated in the following chapter.

Chapter 6. Here we present the empirical study that we performed in order to validate the motivation detection knowledge elicited in chapter 5. In this study we presented participants with an instructional interaction context and asked them to rate the rules that could be applied under those conditions. By doing this, we found which rules from those elicited in the previous study were generally accepted as valid by teachers, and which ones were not.

Chapter 7. In this chapter we present the overall structure of MOODS, a tutoring system simulation developed to test the knowledge elicited via the previous empirical studies.

Chapter 8. It is here that we describe an example interaction with MOODS. MOODS is not an actual tutoring system, but it allows us to quickly test and evaluate the motivation diagnosis knowledge implemented in it by simulating an instructional situation and querying the system about the likely changes in the motivational state of a hypothetical student.

Chapter 9. In this chapter we present the evaluation of the motivation detection techniques implemented in MOODS. In order to perform this evaluation, we performed another study, in which we developed two different instructional simulations, each consisting of six simulated instructional units, and for each instructional unit we asked the study participants to infer the motivational state of a hypothetical student. Then we compared the values inferred by the human teachers to those inferred by MOODS.

The results of the evaluation were encouraging as they seemed to indicate among other things that:

1. The factors used in MOODS to describe an instructional setting convey enough information to make valid inferences about the motivational state of the student.

2. The model used in MOODS to represent the motivational state of the student is considered by teachers important information in deciding the next instructional step to take.
3. The motivation diagnosis knowledge of MOODS seems to be valid in at least some instructional interactions.

Chapter 10. We end up in this chapter by drawing some final conclusions and pointing to some possible further work.

Chapter 2

Background and Related Work

2.1 Introduction

The work relevant to our thesis suffers from what seems to be a paradox: there is a huge amount of it and at the same time there is very little of it. On the one hand, we have the vast psychological literature on theories of emotions, motivation, attitudes, etc. And on the other hand we have the recently fast growing, but still very limited, work on trying to apply these theories to tutoring systems.

This recent interest in affective issues in tutoring systems comes in a context of renewed interest in the role that emotions play in human intelligence and how this affects AI in general, and which is reflected in an increasing number of conferences covering these topics (e.g. Affect in Interactions, 1999; Attitude, Personality and Emotions in User-Adapted Interaction, 2001; Emotion-based Agent Architectures, 1999; Emotion in HCI, 1999; Emotional and Intelligent, 1998; Socially Intelligent Agents, 2000).

This interest in the role of emotions in AI is itself motivated by research in the cognitive sciences arguing that emotions are not detrimental, but rather indispensable, for human intelligence. In one of the most influential recent books covering these issues, Damasio (1996) argues that reason may not be as pure as is commonly held, and that emotions and feelings play an important role in it for better or for worse, suggesting "that certain aspects of the process of emotion and feeling are indispensable for rationality." (Damasio, 1996, pp. xiv–xv).

Although similar arguments are widely embraced and often quoted, questions about their validity have also been raised. Thus, Sloman and Croucher (1981) argue that certain kinds of emotions will be side effects of mechanisms designed to overcome resource limits in intelligent systems. Sloman (1999a) has also argued that "the fact that tertiary emotions are closely related to and depend on mechanisms which are important for higher level management of thought processes doesn't mean that intelligence depends on tertiary emotions, though it does lead to the prediction that some forms of damage will disrupt both".

Regardless of the indispensability or not of emotions for natural intelligence, this debate is also helping increase the interest in research on affective issues in the field of AI-ED, something that has been long overlooked, and which could bring great benefits to the field. But all the related work so far has been, by necessity, very tentative and preliminary, as our current understanding of the influence of affective phenomena in learning is still very limited and too general for a detailed computational implementation of a tutoring system. The influencing factors are many, many of the issues are poorly understood, and many of the theories are vague and/or contradictory and difficult to formalise.

Therefore, the gap between the theoretical work and the practical implementations of tutoring systems is still too big, but an overview of the former can help us understand the difficulties, implications and usefulness of the latter. Thus, on one hand this chapter deals (though, due to the vastness of the subject, far from comprehensively) with our understanding of affective phenomena in humans: What are emotions? What is motivation? How do they affect learning?, etc. On the other hand, we review the work concerned with benefiting from this knowledge in our interactions with computers, and more concretely, in our interactions with instructional systems.

As in many other areas of AI, two different goals can be distinguished in relevant research:

1. to improve the ability of computers to serve humans; and

2. to contribute to our understanding of the concepts at hand (in our case student motivation).

In this thesis the emphasis is on the former goal, and this is reflected in the review presented in this chapter. We start in section 2.2 by untangling the often confusing terms emotion, affect and motivation. In section 2.3 we review work on what is called “Affective Computing”, or computing that relates to emotions. This section does not include work on tutoring systems, but some of the issues and methods there presented could be of direct relevance to developing affective tutoring systems. In section 2.4 we look in more detail at the theoretical aspects of motivation in Education, to go then in section 2.5 to review the relevant research on developing actual “Affective Tutors”. We end by presenting some conclusions in section 2.6, that will lead us to the following chapter 3, where a detailed specification of the work presented in this thesis is given.

2.2 The terminological forest

Throughout this thesis, terms such as ‘emotions’, ‘affect’ and ‘motivation’ are used frequently. Before we proceed any further, we should briefly clarify what is meant by each of them. As Parkinson and Colman (1995) remind us, one of the most widely accepted ways of classifying human mental functions “defines three separate areas of cognition (thinking), affect (feeling), and conation (willing). Emotion is one of the most important and thoroughly explored forms of affect, and motivation is essentially just a new name for conation [. . .]” (Parkinson and Colman, 1995, p. xi).

For this thesis we will treat ‘affective’ as a broad term referring to anything pertaining to emotions or not cognitive, following previous work on what it is usually called ‘Affective Computing’ (e.g. Issroff, 1996; Picard, 1997). But the terms ‘emotion’ and ‘motivation’ need a more careful and detailed definition. There are obvious links between both, as emotions are often precursor of motivational phenomena, and they influence the way we act towards our environment, but they refer to different concepts. As Parkinson and Colman put it, “Emotion and motivation both depend on the relationship between the organism and its environment. In the case of emotion, the emphasis is on the evaluative aspect of this relationship: how the situation makes the person feel; in the case of motivation, it is how the individual acts with respect to the situation that is of interest [. . .]” (Parkinson and Colman, 1995, p. xi). We deal in more detail with these two concepts in sections 2.2.1 and 2.2.2 respectively.

2.2.1 Emotions

To live is to feel, to experience strong emotions.
Stendhal

Emotions are crucial in human life. They are what gives flavour to our lives, but despite this, we still do not know the answers to basic questions in emotion theory, such as “what are emotions?”, “what causes them” and “why do we have them?”. For example, Picard (1997) reminds us that nearly a hundred definitions of emotions have been categorised (Kleinginna and Kleinginna, 1981, cited in (Picard, 1997)), and that there are still many open questions in the theory of emotions (see (Lazarus, 1991, cited in (Picard, 1997)) for a list of twelve of them).

A modern explanation of the term emotion is given by Parkinson (1995, p. 19): “An emotion is a relatively short-term, evaluative state focused on a particular intentional object (a person, an event, or a state of affairs). Good examples are anger, fear, love, and hate.”

But our understanding of emotions has changed greatly in recent years. At the beginning of the twentieth century the James-Lange theory of emotion dominated psychological thinking. This theory’s precursor was William James who postulated that the signs of emotions (facial expressions, etc.) are not the result of an emotion, but rather the emotion itself. After W. B. Cannon’s critique in the 1920s, noting that emotions were still present when the viscera were isolated from the central nervous system, and the years in which behaviourism was dominant, the strongest opposition to this theory came with the renewed concern on cognitive aspects of emotion from cognitive theories.

Stanley Schachter postulated that only a general state of visceral arousal was necessary for the experience of emotion and that the quality of the emotion depended on cognitive, perceptual evaluations of the external world and the internal-state. From this view, new questions about the role of emotions arose, changing our understanding of emotions as interfering with rationality, and prompting the argument (as we have mentioned in section 2.1) that emotions could play an important role on how certain situations and experiences are 'processed' (Mandler, 1987).

Another question which is still greatly debated in emotion theory is which emotions there are (Picard, 1995). Can we find a set of discrete basic emotions or do we have to limit ourselves to continuous dimensions? Some authors argue for basic emotions, the most common being: fear, anger, sadness, and joy. Other authors refer to continuous dimensions, the most common being: arousal (calm/excited), valence (negative/positive) and attention (internal/external source of emotion).

2.2.2 Motivation

Just as for the term 'emotion', there is also plenty of literature about motivation, and many definitions can be found, but generally the term motivation is used to express the disposition of someone to act in a certain way. Thus Weiner writes: "Motivation is the study of the determinants of thought and action—it addresses *why* behavior is initiated, persists, and stops, as well as what choices are made." (Weiner, 1992, p. 17, (emphasis in original)), and according to him there are two main types of motivational theories:

1. *mechanistic theories*; and
2. theories based on a *cognitive approach*.

Mechanistic theories are based on the idea that a human works as a machine and assume that their models of motivation are based on needs, drives and instincts. On the other hand, cognitive theories are based on the idea that humans have choice over the way in which they behave and, therefore, have control over their actions, and assume that their models of motivation are based on thoughts and beliefs (Williams and Burden, 1997b, p. 119).

But the classification of theories of motivation into *mechanistic* and *cognitive* theories is not clear-cut, and some theories share ideas from both approaches. There are many different theories along this spectrum (del Soldato (1994) cites seven: psychoanalytic theory, drive theory, field theory, resultant valence theory, social learning theory, achievement theory and attribution theory). However, the review of each of them is outside the scope of this thesis. Alternatively, some of the basic ideas behind some of these theories will be reviewed, in order to understand the progress made in our understanding of human motivation.

As Williams and Burden (1997b) explain, the early psychological views on motivation (mainly mechanistic theories) were greatly influenced by Behaviourism, and they were mostly based on experiments with animals. These early theories explain motivation with reference to concepts such as biological needs, rewards, and behaviour reinforcement, where rewards are seen as the most effective way to promote learning, and also to increase motivation.

Some of the problems of these early psychological views on motivation are that "they are based on the fundamental principle of *homeostasis*, i.e. they assume that animals and humans prefer not to be in a state of arousal and are constantly seeking to be in a more settled state" (Williams and Burden, 1997b, p. 114). Similarly, they perceive human behaviour as being controlled by external forces and not by our own choices.

The first theories that explain motivation in terms of cognitive processes, although still having similarities to mechanistic theories, are expectancy-value theories. The basic idea behind them is that people 'choose' their behaviour in order to achieve goals of high value. But the choice is considered sometimes, as in achievement theory, to be largely unconscious.

The last step for the development of truly cognitive approaches to motivation is, then, the differentiation of two types of human actions: those under our conscious control and those outside our conscious control. Taking this view, a cognitive approach to motivation "centres around individuals making decisions about their own actions as opposed to being at the mercy of external forces over which they have no control." (Williams and Burden, 1997b, p. 119).

But the theories overviewed above do not necessarily apply well in the field of Education. As Weiner (1984, p. 15) says: "It surely seems unlikely that much of classroom behavior is governed by the sexual and aggressive instincts [...], so the psychoanalytic approach offers relatively little theoretical help. In a similar manner, [theories that focus] on the reduction of biological needs and the survival relevance of behavior, also [are] far removed from classroom concerns." Thus, theories of student motivation present special problems and require departure from prior theories of motivation. An overview of theories of motivation in education is given in section 2.4. These theories will be of particular interest when we look at the relevant work to create 'affective tutors' in section 2.5. But before that, we review in the following section work on what has been called 'Affective Computing'.

2.3 Affective Computing

The question is not whether intelligent machines can have emotions,
but whether machines can be intelligent without any emotions.
The Society of Mind, (Minsky, 1985)

Before we turn to reviewing the work related to instructional systems in section 2.5, we overview in this section the more generic work on what is usually called 'Affective Computing'. According to Picard (the coiner of the term), 'Affective Computing' is "computing that relates to, arises from, or deliberately influences emotions." (Picard, 1997, p. 3).

Despite its different emphasis, affective computing is of direct relevance to our research on affective tutors. One of the main goals of an affective tutor should be to create an instructional situation in which the student is motivated to learn (i.e. he is willing to dedicate some intellectual and/or physical effort in order to learn the topic being taught). But as we have seen in section 2.2, emotions and motivation are highly related, the former being often the precursor of motivational phenomena. Thus, much of the work on detecting and influencing emotions could be applied to research on affective tutors, although the emphasis on which factors to detect and influence will be different, given the peculiarities of theories of motivation in education.

There is a broad range of different research issues on affective computing. *Grosso modo* we can divide affective computing into three big areas:

1. emotion detection

2. emotion synthesis (or simulation), and

3. generation of artificial emotions.

We review some of the research in each of these areas in the following sections.

2.3.1 Emotion detection

Picard (1997) argues that a hallmark of intelligent computers will be their ability to recognise emotions. But an accurate detection of emotions by computers (if at all possible) is still a farfetched goal. After all, not even humans can recognise someone else's emotions 100% accurately. In most cases, this ability to read other people's emotions—empathy—is taken for granted, although not everybody shows the same proficiency at it. Robert Rosenthal and his students measured this ability to empathise with others, by using a test of empathy with over seven thousand people. The test—or Profile of Nonverbal Sensitivity (PONS)— consisted of a series of videotapes of a young woman expressing different feelings (having the words muffled), which were showed to the viewers after blanking out one or more channels of nonverbal communication. The study showed that people differ in their ability to empathise and also that some of the benefits of being able to empathise through non-verbal cues include being better adjusted emotionally, or being more popular (Rosenthal *et al.*, 1977, cited in (Goleman, 1996)).

But as the previous paragraph suggests, the cues available for emotion detection are many. Information for emotion detection is not only verbal. In fact, it is often argued (e.g. Goleman, 1996) that the largest part of an emotional messages is nonverbal, and it is extraordinary how much information is conveyed through other means (the eyes, the hands, posture, body rhythms, smell, touch, etc.), some of which we are not even consciously aware of (Davis, 1976). But it has also been suggested "that for the emotions that matter most to humans the *primary* and most natural mode of expression is linguistic." (Sloman, 1999b, p. 5).

Regardless of which medium is the primary one to convey emotions, a computer could use information from any of these communication 'channels', or ideally, a combination of all of them plus reasoning about emotion-generating situations to detect its user's emotional state. But given our limited understanding of emotions, most of the work so far in emotion detection by computers has focused in a particular information 'channel'. We consider some of this work below.

Recognising facial expressions. As Negroponte (1996, p. 129) put it: "Your face is, in effect, your display device, and your computer should be able to read it, which requires the recognition of your face and its unique expressions". Based on the theory called Facial Action Coding system (FACS) developed by Paul Ekman and his colleagues (Ekman and Friesen, 1977; Ekman and Rosenberg, 1997), Essa and Pentland (1997, cited in (Picard, 1997)) developed a system that can recognise facial expressions from video, showing an accuracy of 98% in recognising (although not in real time) six deliberately made facial expressions for a group of eight people. Other different models for facial expression recognition using pattern recognition have also been developed (e.g. Lisetti and Rumelhart, 1998; Yacoob and Davis, 1996).

Recognising emotions in speech. Recognising emotions in speech is a very difficult task even for humans, who can, on average, recognise affect on neutral speech with about 60% reliability (Scherer, 1981, cited in (Picard, 1997)). Despite this, based on findings on human vocal emotions (such as that a speaker in anger will generally talk faster and louder (Murray and Arnott, 1993, cited in (Picard, 1997))), models can be created that can be implemented in computer systems to recognise emotions. Some preliminary steps have been given in this direction (e.g. Roy and Pentland, 1996; Tosa and Nakatsu, 1996). In the case of Roy and Pentland (1996), they studied the possibility for a computer of discriminating between approving or disapproving sentences, obtaining similar classification accuracy than humans: 65% to 88% for speaker-dependent, text-independent classification.

Physiological Pattern Recognition. Some work has also been done in using physiological data to recognise emotions. By using a Prototype Sensing System with four sensors[1], Elias Vyzas (described in Picard, 1997, pp. 185–188) performed an experiment in which physiological signals were gathered from an actress expressing eight different emotions repeatedly each day, during a period of twenty days. In order to

[1]Galvanic Skin Response (GSR) Sensor, Blood Volume Pulse (BVP) sensor, Respiration sensor and Electromyogram (EMG) sensor. A brief description of the Prototype Sensing System can be found at the World Wide Web: `http://vismod.www.media.mit.edu/vismod/demos/affect/AC_research/sensing.html` (March 3, 2000)

study the possibility of discriminating between different emotions, all triplets of emotions were tried. The best discrimination rates were obtained with two triplets: anger, grief and reverence; and anger, joy and reverence. In both cases, a classification accuracy of 83% was obtained after applying classic pattern recognition tools (Picard, 1997).

Linguistic approaches. As mentioned earlier, some authors (e.g. Sloman, 1999a) have argued that for the most important emotions to humans, the main mode of expression is linguistic. Even Picard, whose work focuses mainly on the use of the bodily expressions of emotions says "[...] emotions are still communicated through the written world. This power and importance of influencing emotions through language was a primary tenet of Aristotle's Rethoric" (Picard, 1995, p. 4).

But given the inherent difficulties of understanding language, progress on detecting affective information through language has been very small. Although, as it has been argued, we will *first* need to understand emotions in order to understand dialogues (Dyer, 1983; Elliott, 1992). One of the few attempts in this direction is the work by Ortony *et al.* (1988), whose "main aim is to derive a computationally tractable calculus of the kind from which a language-understanding program would be able to derive [affective] inferences" (Johnson-Laird and Oatley, 1988, p. 4).

Other. There has also been some interesting research in trying to handle the uncertainty of the emotional state of a user by using probabilistic models. For example, Conati and Zhou (2002) make use of Dynamic Decision Networks (DNNs) (an extension of Bayesian networks) to model affective characteristics of a student during an interaction with an educational game.

2.3.2 Emotion synthesis (or simulation)

Emotion synthesis, as used in this thesis, focuses solely on the simulation of emotions, leaving the issue of a computer 'having' emotions for section 2.3.3. Whether an accurate synthesis of emotions can be obtained without 'having' emotions is disputed, but emotion synthesis could bring real benefits to computer users. Reeves and Nass

(1998) show how people act socially towards machines despite their 'coldness' and lack of affective knowledge and empathy. Machines that are able to show a degree of emotional life could perhaps create a better interaction between man and machine.

The simulation of emotions is not something totally new. For instance, Macintosh[2] computers have always welcomed the user with an artificial smile at the beginning of each working session. The possible uses of this type of research are many (see, for instance (Picard, 1995) where more than 50 possible applications of Affective Computing are given), from practical applications to entertainment. It is mainly in this last area where research into emotion synthesis is most popular (as seen with the commercial successes of the Tamagochi virtual-pets or the artificial dog AIBO[3] from Sony[4]).

But we are more interested in research which based on studies of human emotions makes use of emotion models in order to help the computer to better synthesise emotions. Some examples of this type of research have already been mentioned in section 2.3.1. Thus, the work by Essa and Pentland (1997), although mainly focused on recognising facial expressions, is based on representing facial motion dynamics, and makes use of knowledge about facial shape and muscles, and therefore it can be used to synthesise facial expressions (Essa, 1995, cited in (Picard, 1997)).

In what concerns synthesis of speech, Cahn (1990, cited in (Picard, 1997)) developed the "Affect Editor", a program to generate speech with a desired affect. By using seventeen parameters (for pitch, timing, voice quality and articulation) the program was able to synthesise six types of affect in speech: scared, angry, sad, glad, disgusted, and surprised. To test her model, five neutral sentences were synthesised with the six emotion categories, and listeners were asked to choose which emotion they thought it was expressed. Correct classification was done around 50% of the time, except for sadness, which was correctly classified 91% of the time.

There is also some interesting preliminary work on combining speech and facial expressions. Binsted and Luke (1999) describe early work on a system whose goal is to generate appropriate affective speech and facial expressions, to be used on an

---

[2]Macintosh™ is a trademark of Apple Computer, Inc.

[3]Information regarding AIBO is available at the World Wide Web: `http://www.world.sony.com/Electronics/aibo/index.html` (March 15, 2000).

[4]AIBO™ and Sony™ are trademarks of Sony Corporation, Tokyo.

animated talking head system.

The focus of some other work has been mainly on understanding human emotions, but providing interesting models for emotion synthesis, e.g. the PARRY program by Colby (1981) and the Affective Reasoner by Elliott (1992). The well known program PARRY has become a classic in AI literature. It simulates the behaviour of a schizophrenic paranoid who has three affective responses (fear, anger and mistrust) that can be triggered in different situations if malevolence is detected from the input in a simulated psychiatric interview. Similar to the approach taken for the work presented in this thesis, the detection of malevolence and the selection of responses is determined by the values of a small number of variables in the program.

Elliott (1992) focused mainly on designing a pragmatic and rich computer representation of emotions, congenial with emotion theories. He implemented a multi-agent world populated by several agents who can make inferences on emotional states of other agents by using representations of twenty-four emotion types. This platform allowed for the reasoning about emotions, supporting the testing of emotion theories.

2.3.3 Generation of artificial emotions

To 'successfully simulate' intelligence is not the same as being intelligent. This old and controversial debate in AI can easily be translated to the issue of emotions. Would being able to successfully detect and synthesise emotions necessarily imply 'having' emotions? The implications of this issue are very different to those of simple simulation of emotions, but this philosophical debate is of no direct relevance to our thesis (and therefore we will not discuss it here), except perhaps for one of its implications. Goleman has argued that the ability to empathise "builds on self-awareness; the more open we are to our own emotions, the more skilled we will be in reading feelings" (Goleman, 1996, p. 96). Will this imply that a computer should *have* emotions in order to be able to empathise with its user?

2.3.4 Conclusion

In this section we have overviewed some of the research on endowing computers with a degree of affect (be it to detect its users' emotions or to communicate its own (simulated) emotional states). Some of the work seems to offer promising results, and the growing interest in recent years in this area promises new and interesting research possibilities. Despite this, and due to the enormous difficulties of the task at hand, we believe that the issue of detecting human emotions by a computer should be approached as an attempt to detect an 'emotion direction', a tendency, or to discriminate between a limited number of emotions. To expect a computer to be able to recognise accurately all types of emotions at any time is simply unrealistic and unfeasible with our current understanding of emotions.

This research also raises a number of very controversial sociological and ethical questions. The social implications of AI are many (see, for instance, (Whitby, 1996)), and some authors (e.g. (Warwick, 1998)) have aired an extremely alarmist opinion of the dangers that technological advances are creating in our society. The amount of implications to our society (and of alarmist viewpoints) will no doubt increase proportionally to the degree of success of creating 'Affective technology', but the benefits are also many. We believe that the incorporation of emotional issues into the research agenda of AI could improve the quality of our interactions with machines and also help us to further understand human emotions.

2.4 Motivation in Education

Before we review the work concerning "Affective Tutors" in section 2.5, in this section we review some of the theoretical work done in understanding the role of motivation in learning. The importance of keeping students motivated can be felt by anyone who has experience with practical teaching. As Goleman reminds us, "The extent to which emotional upsets can interfere with mental life is no news to teachers. Students who are anxious, angry, or depressed don't learn; people who are caught in these states do not take in information efficiently or deal with it well." (Goleman, 1996, p. 78). But perhaps what is not so intuitively recognised is the extent to which this is dealt in instructional situations. As Lepper *et al.* (1993) report after studying the motivational techniques of expert tutors in elementary school mathematics activities, human tutors seem to devote at least as much time and attention to the achievement of affective and motivational goals as to cognitive and informational goals.

But even if they are well-intentioned, the efforts of teachers to motivate their students are not always successful, probably due to lack of training and to our lack of deep understanding of all the issues involved. To illustrate some of the problems of some instructional approaches[5], we point to four faulty assumptions underlying school motivation practices, highlighted by Clifford (1990):

1. "... with regard to learning, students are passive and must be given detailed directives if learning is to occur"
2. "... learning is aversive and engaged in only when mandated by external forces."
3. "... student success, operationally defined as high performance, is the best of all motivation potions!"
4. "... error-making and failure are necessarily detrimental and reduce self-esteem, motivation, and learning"

(Clifford, 1990, pp. 61–62)

What, then, is the correct approach to motivating a student? Although a definitive answer to this question cannot be given, we look at this issue in the following sections. In section 2.4.1 we review briefly Keller's theory of motivation in Education to help us understand the issues at hand. In section 2.4.2 we look in more detail at what the factors that influence students' motivation are, according to a number of theories of motivation in Education.

2.4.1 A theory of motivation in Education

In this section we review Keller's (1979) theory of motivation in Education. But before we review it we should point out the necessary incompleteness of any attempt

[5]Including much of the research on the field of AI-ED.

at explaining the complex issue of motivation, due to the enormity of the task. In Weiner's (1984) words:

1. A Theory of Motivation Must Include the Full Range of Cognitive Processes.
2. A Theory of Motivation Must Include the Full Range of Emotions.
3. A Theory of Motivation Must Explain Rational and Nonrational Actions, Using the Same Concepts for Both.

(Weiner, 1984, p. 15)

The work by Keller is very interesting as he has developed a theory of motivation, performance, and instructional influence (Keller, 1979) and a model for motivational design of instruction based on that theory (Keller, 1983). Keller's (1979) theory (which is graphically summarised in figure 2.1) is a macro-theory, that subsumes important concepts from the study of instruction and learning, and whose purpose is to identify major categories of variables of individual behaviour and of instructional design that are related to individual effort and performance. Thus, Keller sees behaviour as a function of the person and the environment and the theory describes the influence of these two factors on three categories of responses: effort, performance, and consequences.

Effort and performance are clearly distinguished in this theory. Effort refers to whether the individual is engaged in actions aimed at accomplishing the task, while performance means actual accomplishment. The distinction is necessary, as effort is a direct indicator of motivation, while performance is not (although many studies use performance measures as indicators of motivation). Performance is only indirectly related to motivation, and is affected by other factors, such as individual abilities or learning design of the instruction. This can be seen in figure 2.1. According to this model, the effort that an individual puts into a task is influenced by three broad variables:

1. Motives (values), which refer to how individual needs and beliefs relate to choices of action.

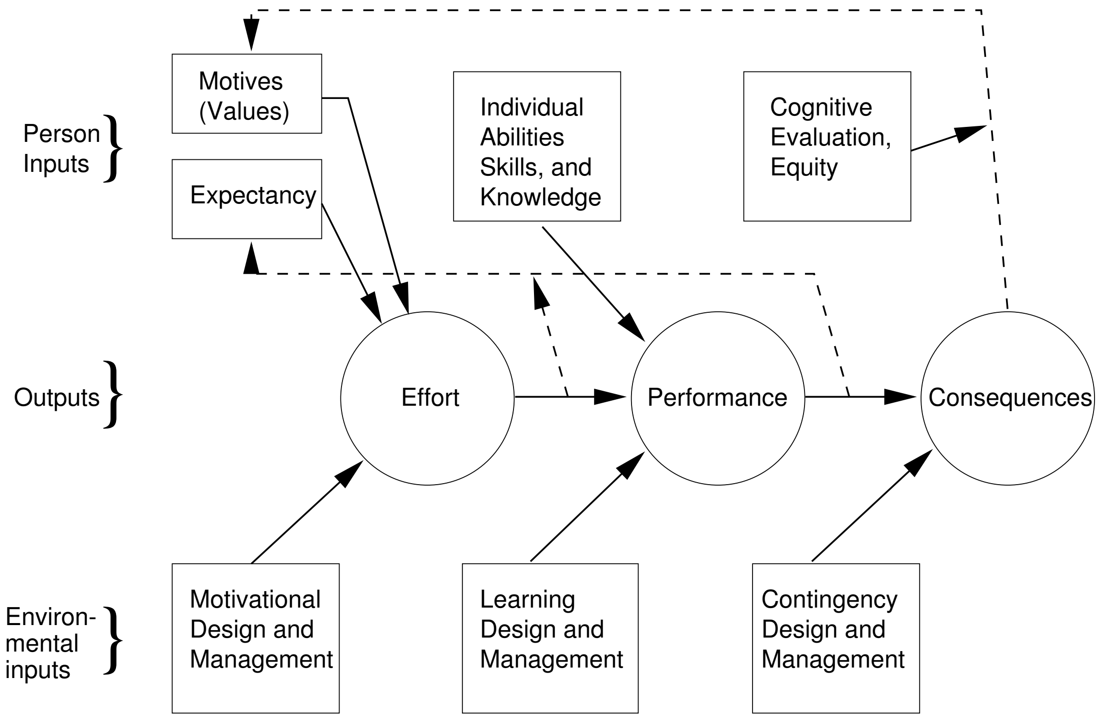


Figure 2.1: A model of motivation, performance, and instructional influence (reproduced from Keller, 1983, p. 392).

2. Expectancy, which is concerned with how personal expectancies for success or failure affect behaviour.

3. The motivational design and management of instruction.

On the other hand, performance is influenced by:

1. Individual abilities, skills, and knowledge.

2. Learning design and management.

3. The effort the individual puts into the task.

The third category of responses in this theory is 'Consequences', which refer to the intrinsic and extrinsic outcomes that accrue to an individual (e.g. emotional responses, social rewards, material objects). These consequences are related to the performance and to the (environmental variable) “Contingency Design and Management”. These consequences play an important role in motivation as they feed back (through cognitive evaluation) into the motives and values of the individual. For example, a positive performance followed by an external reward (such as cash) may, as a consequence, influence the value that the individual places on the activity.

2.4.2 How to influence student’s motivation?

As can be seen from Keller’s (1979) theory reviewed in section 2.4.1, there is a vast number of influences on a student’s motivation. But a study of which are the most important factors and how they affect motivation can help to design more motivating instruction.

Lepper and Chabay (1988) looked at human tutoring sessions in one-to-one settings, finding that tutors seem to have three main motivational goals:

1. “keep pupils from becoming so discouraged, frustrated, or alienated that they give up on the task at hand”

2. “encourage in their students high levels of attention and effort”

3. improve the conditions “that promote intrinsic motivations for learning.”

(Lepper and Chabay, 1988, p. 246)

But how do tutors accomplish these goals? Which are the variables that they can manipulate during the instruction? In table 2.1, we present the most important motivational goals of instruction, as suggested by Malone and Lepper (1987). From their work, we also summarise the factors that influence these goals and some suggestions as to how to achieve the desired motivational goals.

Table 2.1 offers a taxonomy from which guidelines for instructional design can be obtained and it could also be used to evaluate the motivational ‘fitness’ of the research reviewed on the following section regarding ‘Affective Tutors’. But as Malone and Lepper (1987) remind us (regarding their proposed guidelines), these suggestions should not be treated as a prescription, but rather as “a way of guiding and sharpening

Goals	Influences	Suggestions
Challenge	Level of difficulty Goals of task Performance feedback Learner’s sense of self-esteem	Provide intermediate level of difficulty and challenge Explicit, proximal and personally meaningful goals Goal attainment uncertain Performance feedback that will engage and enhance the self-esteem (frequent, clear, constructive, encouraging)
Curiosity	Informational complexity Discrepancy or incongruity from present expectations and knowledge	
Sensory curiosity	Light, sound, or other sensory stimuli of an environment	
Cognitive curiosity	Higher-level cognitive structures	Making people believe that their existing knowledge structures are not well-formed (ie. they lack completeness, consistency or parsimony)
Sense of control	Range of outcomes provided by environment Probability of the person to influence those outcomes	Provide contingency (individualisation) Provide choices (apparent and salient responsiveness) Students’ actions should have ”powerful effects”
Fantasy	The emotional needs they help to satisfy (power, success, fame, fortune, etc.) Identification with imaginary characters (based on a) perceived similarity, b) admiration, and c) salience of that character’s perspective	Create different fantasy contexts which users can choose from.

Table 2.1: Factors influencing student’s motivation (Malone and Lepper, 1987).

intuitions and aesthetic sensitivity, not a way of replacing them" (Malone and Lepper, 1987, p. 249).

2.5 Affective Tutors

In section 2.3 we have overviewed some of the work concerning how computers could detect user's emotions and/or how they could show (simulated or real) emotions. The focus of this section (and of the main focus of this dissertation) is on instructional systems that take into account more than just cognitive aspects of instruction. We call these 'Affective Tutors' following a similar terminology to the name 'Affective Computing', but the emphasis of the two sections is very different. In section 2.3 we have overviewed general issues and techniques for introducing computers to the realm of affective phenomena. In this section the emphasis is on how to use these techniques towards a very specific goal: to create instructional systems that engage the student, that motivate him. As we have seen in section 2.4, motivational issues are a primary concern for human tutors. Despite this, there has been little research in the field of ITSs that has paid attention to them.

One of the first suggestions of endowing computer tutors with a degree of empathy was made by Lepper and Chabay (1988). They argued that motivational components are as important as cognitive components in tutoring strategies, and that important benefits would arise from considering techniques for creating computer tutors that have an ability to empathise. They also give some suggestions of what should be added to an ITS so that it can empathise with students' feelings:

1. General social knowledge or "some general rules concerning the appropriateness of different sorts of social and motivational remarks in various situations" (p. 251).

2. Specific background knowledge about the individual student.

3. A component to offer choices to the student and to analyse his responses (for instance, if he would like help, an easier problem, etc.).

As we can see, the issues involved in creating an affective tutor are many, and so far we have assumed mainly individualised instruction. Although the focus of this thesis is on individualised instructional settings, we should bear in mind that if we look at other types of instruction, there are many other issues that we should pay attention to, as we can see from the work of Issroff (1996). Issroff performed a detailed study of the affective influences of Computer-Supported Collaborative settings, based on observations that in different instructional settings (i.e. cooperative, individualised and competitive), the required motivational techniques are different. This is so, partly due to the different values of participants in each of these situations. Thus, for example, an important aspect for the participants in the individualised setting is one's performance over time, whereas in the competitive situation the important aspect is social comparison information and in the cooperative situation is group performance.

Despite this complex picture, there have been some attempts to create 'affective tutors', with varying degrees of attention to motivation in Education theories and with different depth. We review this work in the following sections.

2.5.1 Implicitly affective tutors

Before we continue any further, this section helps to highlight what this thesis is *not* about. This thesis is not about finding general motivating techniques that can be incorporated into tutoring systems. Many tutoring systems attempt to motivate the student by using multimedia, games, etc. This approach seems to be based on the idea that it is possible to create instruction that is motivating *per se*. While this approach can probably yield the development of useful tutoring systems, the emphasis in our work is to explore how a tutoring system could be individualised in the affective realm. In this section, then, we just present some examples of how this idea of motivating instruction *per se* permeates some of the research on tutoring systems.

As Keller (1983) mentions, many instructional technologists assume that if the instruction is of good quality motivation will follow, but this is not always the case, since "we can have courses that are of demonstrably better quality with respect to the learning objectives, but less appealing than the comparison groups." (Keller, 1983, p. 388). Similarly, Mark and Greer (1993) remind us that it is frequently argued that computers are intrinsically motivating or that their use may foster self-esteem. This is particularly so when new or uncommon techniques are used (for example, regarding the use of 3D techniques Bares *et al.* (1998) say: "Properly designed, 3D learning environments that blur the distinction between education and entertainment could produce learning experiences that are intrinsically motivating and solidly grounded in problem solving." (Bares *et al.*, 1998, p. 77)). But it seems obvious from the literature on motivation in Education that one of the main factors influencing motivation is curiosity, which no doubt will fade out if the only strategy for instruction is to use a particular teaching technique or environment. We suspect that the original motivation will quickly fade away as the particular setting becomes habitual.

Another issue which is currently being investigated by many researchers in the area of ITSs is the effects of collaborative or cooperative student models. The positive motivational effect of this type of student models has also been suggested by some authors, arguing that the student would perceive that he controls his learning (e.g. Beck *et al.*, 1997). Similarly, many of the new techniques and approaches being researched in the area of ITSs nowadays (such as companion agents, learning by disturbing, emotional agents, etc.) suffer from the same limitation. On their own they can be amusing and engaging (probably due to their novelty), but to attach them an enduring intrinsically motivating value does not, we believe, approach the issue of creating motivating instructional tutors in the best possible way. This is not to say that this type of research is not useful. Some of this work (for instance, work on emotive communication in pedagogical agents, e.g. (Lester *et al.*, 1997, 1999), or the work on multimedia instruction based on motivational principles by Stoney and Oliver (1998)) can produce very interesting and encouraging results.

We believe that we should consider this type of research as a starting point, as suggestions of useful techniques for motivating the student, but ideally we should be able to make use of theories and/or empirical studies that could tell us under which circumstances these techniques can have the best motivational influence for a particular student. Otherwise, we believe that we will very often find that very interesting techniques are not used as effectively as they could be (see for example the work by Hietala and Niemirepo (1997), where agent companions with very interesting human-like abilities to make mistakes and to learn slowly were found, not surprisingly, to be liked by some learners but not by others).

The main concern in this thesis is the development of individualised motivating instruction, and for this a crucial issue is affective 'detection'. Work on instructional tutors that have incorporated not only motivational strategies but also some sort of affective detection is reviewed in section 2.5.3. Before that, we look in section 2.5.2 at some other relevant work for affective tutors, but whose focus was not on creating a tutoring system *per se*.

2.5.2 Relevant issues for affective tutors

Before we turn to review work on creating a complete affective tutor, we review in this section some relevant research to this task. The work reviewed here does not focus necessarily on affective tutors, but it is of direct relevance to our research. We review relevant work on detecting motivation, modelling it in a computer, planning motivational instruction, and eliciting affective knowledge, or how to obtain the detailed knowledge required for an affective tutor implementation in a computer.

Motivation diagnosis

As we have argued previously, we believe that the issue of detecting the student's motivational state is crucial for developing successful affective tutors, but to do this is not straightforward. As Clifford (1990) reminds us, there are many school indices such as grade-point average or aptitude scores, but there are no indices for motivation (at least not indices ready for the classroom use. For some motivational instruments used in research, see Vidler, 1977). Therefore, we should consider alternative methods for detecting motivation. As we argued in (de Vicente and Pain, 1998), there are many possibilities for approaching the issue of detecting the motivational state of a student, some of which have already been tried out. We summarise below some of these attempts, classifying them according to the type of information used for the diagnosis.

Questionnaires. Questionnaires are a widely used approach to measure motivation in classrooms. Questionnaires already exist to measure motivation in certain settings.

For instance, see the work by Gardner (1985) who devised an Attitude/Motivation Test Battery (AMTB), which consists of a series of self-report questionnaires, in order to calculate an Attitude Motivation Index (AMI). There also exist guidelines to developing questionnaires to measure motivation in particular settings (e.g. O'Bryen (1996) gives an introduction to the use and development of questionnaires to assess motivation in second language classrooms). But questionnaires to measure attitudes pose extra difficulties and there is ongoing debate of how accurate a tool a questionnaire can be (see Oppenheim, 1992) . Therefore, the use of questionnaires is not as simple as it may seem. Besides, questionnaires are static, but ideally we would like to measure not only permanent characteristics, but also the changing motivational state of the student during the interaction.

Questionnaires could be useful for collecting information about enduring characteristics of the student at the beginning of an interaction, while using other methods to gather information about more transient characteristics of the student's motivation. It is with this approach that some previous research has used questionnaires in tutoring systems to detect motivational aspects. Thus, Arshad (1990, cited in (del Soldato, 1994)), used questionnaires applied at the beginning of the first interaction in order to model the student's confidence state.

Matsubara and Nagamachi (1996) also used questionnaires[6] at the beginning of the interaction to diagnose several factors influencing motivation, such as: Achievement Motive, Creativity, Sensation Seeking Scale, Extroversion Intention, Work Importance and Centrality, Least Preferred Co-worker and Locus of Control.

Whitelock and Scanlon (1996) used post test questionnaires (consisting of five open ended questions) to assess a number of motivational factors, such as "curiosity, interest, tiredness, boredom and expectation plus the challenge of the task and partnership itself" (p. 276).

Self-report. An interesting approach to motivation diagnosis that can be used for transient motivational characteristics is self-report. An interface could be easily implemented with features that would allow the student to report his subjective reading

[6]Although no precise account or reference to how the questionnaires were created or what they 'look like' is given in the paper.

of the motivational factors to model. For example, as we will see later in chapter 7, each of these factors can be represented by a slider, which can be manipulated by the student during the interaction. Or a visual scale could be used, as Sobral (1992) did. The scale Sobral (1992) used consists of a two-dimensional space in which the two main factors affecting course appeal are placed in orthogonal scales. Thus, the student only has to draw a mark in one point in this space in order to indicate a measure of how appealing he finds the course. The same type of measure was used in the famous research by Lang (1995) on classifying pictures according to their arousal and valence.

The self-report approach is suggested by del Soldato (1994) for the design of the system MORE, but we learn that, unfortunately, the features needed to read the self-report of the student's motivational state were not operational in the first version of the system, and hence we do not have an evaluation of how effective these features can be.

Spensley *et al.* (1990) also mention the use of self-reports by a system developed by them in order to model some short-term states of the student (e.g. bored, confused). Similarly, Issroff (1996) also used student's own perceptions in order to assess affective factors in her work.

The self-report approach, as that of questionnaires, suffers from similar concerns. Both of them use students' perceptions of affective measures, a method which is sometimes criticised because it can perhaps provide inaccurate measurements (see O'Bryen (1996) for a summary of the debate in past research on self-report measures). But self-report has also the main advantage of its ease of elicitation and many authors have supported its use (e.g. Weiner, 1984).

So, it is not clear whether or how the student's motivational state will be affected by having to report about his own motivation. But, if it is not affected significantly, this could probably be one of the easiest ways to diagnose it. As Briggs *et al.* (1996) put it, "confidence judgements are extremely simple to elicit, since users can give subjective ratings more easily than they can offer explanations."

Expert System. Another interesting approach that can be used in an affective tutor is to provide an "expert system"component that could help the system to find the causes of a performance problem. This approach would have the advantage of not being as intrusive as other approaches (such as self-report), but the disadvantage of being less dynamic.

Although their work was not focused on computer tutors, an example of this approach can be seen in the work of Hioe and Campbell (1988). They were concerned with performance problems in the workplace, and they devised a prototype expert system to help the employee's manager or supervisor to find which problems were affecting the employee's performance, with special emphasis on motivational problems. By reviewing different theories of human motivation, and based on the expert's diagnostic processes[7], they classified potential motivation problems into four groups: 1) performance standards and goals; 2) positive and negative outcomes; 3) human relation issues; 4) work itself. Then, they created a set of questions for specific motivational problems in these four groups, and developed an expert system that could ascertain which of these motivational conditions was causing the poor performance.

Sentic Modulation. Sentic modulation refers to "the physical means by which an emotional state is typically expressed" (Picard, 1997, pp. 25), such as voice inflection, facial expression, and posture. Although, to our knowledge, there is no published work on using this approach in computer instruction systems, Picard (1995) reports the interest in the MIT Media Lab of building systems for teaching piano-playing that are able to detect student's expressive timing, etc. As reviewed in section 2.3.1, there has been some work on detecting emotions by these means, which could be used in an affective tutor.

But we should also consider the difficulty of applying these methods to current ITSs, and in the case of physiological data, the possibly negative reaction of students to the use of body sensors. It could happen that despite being a very efficient way of detecting human emotions, its use could be limited to a certain type of application, as a consequence of users' reaction. For example, in an informal survey among 10 students at our University we found—perhaps not surprisingly—that the use of physiological data was considered the most intrusive approach, and the one that they would least willingly use.

[7] The expert being the second author.

Affective knowledge elicitation

As we have already mentioned, generic rules or techniques do not offer an acceptable solution to the issue of creating affective tutors. In table 2.1 we have seen a number of suggestions for creating motivational instruction, but many of them are vague or too general for their implementation in a computerised tutor. For instance, the suggestion number 1 by Lepper and Chabay in page 29, to include "*some general rules concerning the appropriateness of different sorts of social and motivational remarks in various situations*" could hardly be more general. From which sources can we find the required affective knowledge with an appropriate granularity for implementation in a computerised system?

The first place to look for this information is obviously in the theories of motivation in Education (Keller's (1979) theory was summarised in section 2.4.1). The problem with motivational theories is that their usefulness for the creation of an affective tutor is limited, as the following quote reminds us:

> Motivation theories provide at best moderate predictive and explanatory power for learning and achievement under classroom conditions. Because of this, theory-based recommendations to practitioners in the field are often too vague, too contradictory, and too abstract to be really useful.
>
> (Weinert, 1990, p. 91)

Also, as Lepper and Chabay (1988) point out, even if one knew how a human tutor should behave, there would be questions of whether this knowledge would be appropriate for computer-based tutors.

Other uses of affective information

In the previous sections we have reviewed work on creating computerised tutors that are able to empathise with students, but knowledge about motivation in Education can also be applied in other interesting ways. As an example, we can mention the work by Yin *et al.* (1998), in which they present a prototype system in which a classroom is simulated and where there are different students with different personalities. The user of the system (who is training to be a teacher) can try different tutoring options and see how these affect her students.

2.5.3 Affective tutors explicitly

In this section we review the work by del Soldato (1994), which is to date the most relevant work in this area, as it is based on motivation theories, and attempts to encompass all the necessary steps of creating an affective tutor[8].

Concerning the diagnosis or detection of the student's motivational state, she differentiates four sources of information that can be analysed:

1. Questionnaires applied at the beginning of the first interaction.
2. Communication with the student during the interaction. She mentions the possibility of avoiding a natural language interface by providing the student with a set of standard expressions (e.g. "OK", "too difficult", "hint please", etc.).
3. Student's requests for help and perseverance to complete the task.
4. Learners' self-evaluation of their motivational state during the interaction.

(p. 25, edited)

Two systems were developed: a generic instructional planner called MORE ("MOtivational REactive plan"), and a Prolog-debugging tutor application as a 'vehicle' for MORE. In this application, the second and third sources of motivational information from the list above are used, leaving the use of the first and fourth as suggested possibilities for further research.

The motivational information gathered from these sources is used to modify the model of the student's motivation, which is characterised as a set of three numerical variables: effort, confidence and independence. As the focus of her work is the instructional planner, there are several issues concerning the motivational modelling that could be further developed. Among others, she mentions the following limitations of her system in relation to the motivational modelling:

[8]There is also some more recent research that deals explicitly with motivational issues in ITSs, such as (Georgouli, 2002), but this is mostly based on the work of del Soldato (1994), and therefore is not reviewed separately here.

- the lack of communication channels, such as natural language and vision in the interface of her system, make the detection of the student's motivational state very limited;
- the 'quality' of the motivational features could be improved. For instance, the confidence and independence features are changed 'independently', and the existence of possible correlations was not considered.

In order to clarify how the system diagnoses and models the student's motivation, some of the rules used for the modelling of the student's confidence value are shown in table 2.2.

Rule	**Steps**	**Answer content**	**Confidence model**
C3	none	*help request*	decrement by *conf-dec*
Rule	**Problem**	**With/without help**	**Confience model**
C4	*succeeded*	without *help*	increment by *large-conf-inc*
C6	*failed*	without *help*	decrement by *conf-dec*
C7	*failed*	with *help*	decrement by *large-conf-inc*

Table 2.2: Some rules for confidence modelling, from (del Soldato, 1994, Table 4.7, p. 39)

When a student asks for help without trying to perform the task it probably means that his confidence is low, and so the system will use rule C3 to decrement the confidence value by a constant *conf-dec*[9]. Rule C4 reflects the case when the student succeeds in performing the task without help, and the system will increase the student's confidence value by a constant *large-conf-inc*. On the other hand, if the student fails to perform the task without help (rule C6), the decrease of the confidence value will be smaller than if he fails to perform the task when help was provided (rule C7).

The next issue is how to use the information of the motivational model in order to plan the following steps during the instruction, what she calls 'Reaction to Learners'

[9] In MORE, the range for the confidence value was 0–10, *conf-inc* was set as 1, *conf-dec* as -1, *large-conf-inc* as 2, and *large-conf-dec* as -2

Motivational State'. She compares the reaction of a typical domain-based planner with the reaction of a motivational-based planner. As an example, in the case where the student succeeds in performing a task, the typical domain-based planner would suggest a harder problem. On the other hand, a motivational-based planner would take into account other variables, such as effort and confidence. In her formalisation of the motivational tactics, she suggests that by using binary variables[10] for effort and confidence, four different situations arise when the student succeeds in performing a task, as can be seen in table 2.3.

	Confidence	
Effort	**low**	**ok**
little	(prevent disappointment) comment: difficulty-level-promotion next problem: harder	(stimulate challenge) comment: suggest challenge next problem: much harder
large	(increase experience of success) comment: link effort to success next problem: similar	(ideal situation) comment: performance feedback next problem: harder

Table 2.3: Motivational Planner—Teacher's actions when Student succeeds in solving a problem. Adapted from (del Soldato, 1994, Table 3.2, p. 28)

Thus, if the effort has been 'large' and the confidence is 'OK', the planner will suggest a harder problem, coinciding with the suggestion of the typical domain-based planner. But if the effort has been 'little' and the confidence is 'OK', then the planner will suggest a much harder problem. See (del Soldato, 1994, Chapter 3) for the motivational tactics of the planner in the situations when: the student fails in performing the task; the student gives up performing the task; and the student requests help.

Sometimes, as seen above, the domain-based planner will suggest a different action from that suggested by the motivational-based planner. To resolve these conflicts, she also implemented a negotiation planner, that is responsible for deciding "between

[10] In the actual implementation of MORE and the Prolog-debugging tutor, the variables are not binary, but they are basically used as binary for the instruction planning. Their values are in a linear scale, but a threshold value divides this scale in two, basically using the variables as if they were binary.

traversing the domain or increasing the student's motivation", p. 45. The description of the implementation of the different planners would be too extensive to show here, but two of the rules of the negotiation planner are shown in table 2.4.

Rule N1 deals with the situation where the student succeeded in performing the task, but his confidence is 'low'. Then the action suggested by the domain-based planner 'suggest problem type harder' will be deleted and changed by 'suggest problem type similar', in order to follow the tactics suggested by the motivational planner 'increase *experience success*' and 'not *stimulate challenge*'. Rule N2 deals with the situation where the student effort was 'little', but his confidence is 'OK'. In this case 'suggest problem type harder' will be replaced by 'suggest problem type much-harder' and a comment about the new challenge will be provided to the student. In the rules shown here, the motivational planner overrules the actions of the domain-based planner, but this is not always the case, and in some situations the motivational planner will provide a complementary action.

The issue of 'when' to provide 'which' tutorial intervention is also of importance to the negotiation planner, and there are five rules to decide whether to intervene or to skip help. For instance, when the student asks for help the tutor will suggest that the student (following the rule N12) works without help if the suggestions of the motivational planner include the tactic *encourage independence*.

2.6 Conclusions

In this chapter we have briefly reviewed a very large amount of work. We have looked first at terminological issues in order to make clear the concepts of affect, emotion and motivation. While most of the work on affective computing has focused on emotions, we have argued that for tutoring systems the main affective concern is that of motivation. But the issue of motivation, as we have seen in section 2.4.1, is heavily dependent on emotions. Due to this dependency and their similar implementation concerns of 'Affective Computing' systems and 'Affective Tutors', we have reviewed work in both areas.

As we have seen from this background and related work chapter, there is a growing

	DOMAIN-BASED PLAN	**MOTIVATIONAL PLAN**	**NEGOTIATION PLANNER**	
Rule	**Action**	**Tactic**	**Delete action**	**Add action**
N1	suggest *problem type harder*	increase *experience success* not *stimulate challenge*	suggest *problem type harder*	suggest *problem type similar*
N2	suggest *problem type harder*	*stimulate challenge* not *increase confidence*	suggest *problem type harder*	suggest *problem type much-harder* provide *comment content challenge*

Table 2.4: First two rules of the negotiation planner, from (del Soldato, 1994, Table 4.12, p. 46)

interest in these issues in the field of AI-ED, but there is still no clear framework or clear theories on which to base our work. Thus, the work in this area is, by necessity, of a tentative nature, and its evaluation extremely difficult. Our work, which also suffers from these limitations, is summarised in detail in chapter 3.

Chapter 3

Outlining the design of an Affective Tutor

Based on the work reviewed in chapter 2, and specially in section 2.5, we present in this chapter the outline of the design of an ITS with motivation diagnosis features. This design constitutes the basis for the rest of the work in this dissertation. After an introduction in section 3.1, we focus on the design of a motivation model in section 3.2. Then, in section 3.3 we discuss the different techniques that we will consider for the development of an Affective Tutor.

3.1 Motivation modelling

Student modelling is one of the main issues in AI-ED research. It seems obvious that the only way that instruction can be individualised to fit a particular student is by "getting to know" the student. Considerable research has been done in this area (e.g. Greer and McCalla, 1994; Self, 1988a,b; VanLehn, 1988), and in the biennial International Conference on User Modeling (e.g. Bauer *et al.*, 2001; Jameson *et al.*, 1997; Kay, 1999) a large amount of the contributions are regularly devoted to the peculiarities of user modelling in ITSs.

Several diagnostic techniques have been identified and are considered 'core' student modelling techniques (e.g. VanLehn, 1988), but the field also provides opportunities to investigate new techniques and uses of student models, such as the pedagogical issues of building models collaboratively between student and system (e.g. Bull, 1997; Bull *et al.*, 1994).

Self (1988b) reviewed 18 well-known ITSs and identified 20 different uses that have been given to student models. By reviewing this list of student model uses, we see that all of them refer to the modelling of cognitive characteristics of the student. In this section we focus on modelling affective characteristics. Ideally, a tutor should use a combination of these, and base the adaptation of the instruction both on the student's intelligence and motivation.

In section 2.5.2 (and in de Vicente and Pain, 1998) we have reviewed some of the possible approaches that could be used for detecting a student's motivation. As we explain in detail later on, we focus on four of them: questionnaires, self-report, knowledge-based rules and verbal communication. But before we continue, in order to make the following exposition more clear, it is useful to clarify two different student motivation modelling processes: *diagnosing* and *representing*. *Diagnosing* refers to the processes by which information regarding the student is elicited. *Representing* refers to the formalisation and representation of that information in an adequate format.

3.2 *Which* motivational factors to model?

As we have seen in section 2.4, there are many factors that influence the student's motivation. Some of these are environmental while others are personal, but theorists disagree on which these variables are. In order to decide which variables to model we follow a number of suggestions made by Self (1988a, pp. 23–24) on how to "bypass the intractable problem of student modelling":

- Don't diagnose what you can't treat.

- The information needed by the ITS to build the student model should be provided naturally by the student, and not inferred from inadequate data.

- Explicitly link the proposed contents of the student model with specific tutorial actions, ideally supported by educational evidence.

Therefore, we limit our motivation model to broad categories, which arise from the motivational categories summarised in table 2.1, and for which there exist suggestions on how they affect the student's motivation and how they should be treated. How this information is obtained is dealt with in this and subsequent chapters, leaving the issue of how to link this information with tutorial actions for chapter 7.

The model we propose, which can be seen in figure 3.1, is divided into two main categories: trait variables, or 'permanent' characteristics of the student; and state variables, or more 'transient' characteristics. This distinction between traits and states is common in the relevant literature, and help us distinguish between those characteristics of the student that are not likely to change during an instructional interaction, and those characteristics of a transient nature, that are likely to change during an instructional interaction. The information about trait characteristics would allow an Affective Tutor to individualise the instruction based on student prototypes, while the state variables would allow an Affective Tutor a more detailed individualisation based on changes during the interaction with the system.

The actual position of each variable under one of these categories (traits and states) in our model is not necessarily the only possible way of categorizing it. It could be argued that some of the variables could equally be placed under the other category. For example, "confidence" could be interpreted as how confident a student is generally, and it could be placed under the "Traits" category. Similarly, "control" could be interpreted as a more transient characteristic, meaning the degree of control that a student wants for each instructional item. The category where a variable is placed in our model gives an indication of the emphasis given to it in our model. Thus, we consider "confidence" a more transient characteristic, focusing on how confident the student is throughout an instructional interaction. On the other hand, we consider "control" a more permanent characteristic, one that reflects whether a student likes to have control over his learning environment. We believe that the categorization that we propose is in tune with current theories of motivation in Education. We give a definition for each of these variables in table 3.1.

Variable	**Definition**
Control	Refers to the degree of control that the student likes having over the learning situation (i.e. does he like to select which exercises to do, in which order, etc. rather than let the instructor take these decisions?).
Challenge	Refers to the degree that the student enjoys having challenging situations during the instruction (i.e. does he like to try difficult exercises that represent a challenge for him?).
Independence	Refers to the degree that the student prefers to work independently, without asking others for help (i.e. does he prefer to work on his own, even if he finds some difficulties, and try to solve them by himself rather than asking collaboration or help from others?).
Fantasy	Refers to the degree that the student appreciates environments that evoke mental images of physical or social situations not actually present (i.e. does he like the learning materials being embedded in an imaginary context or does he prefer just "the facts"?).
Relevance	Refers to whether the student meets *important personal needs* throughout the learning situation (i.e. does he think that the instruction materials are personally relevant to him?).
Confidence	Refers to the student's belief in being able to perform the task at hand correctly.
Sensory interest	Refers to the amount of curiosity aroused through the interface presentation (i.e. appeal of graphics, sounds, etc.).
Cognitive interest	Refers to curiosity aroused through the cognitive or epistemic characteristics of the task (i.e. regardless of the presentation issues, does the student find the task at hand cognitively appealing?).
Effort	Refers to the degree that the student is exerting himself in order to perform the learning activities.
Satisfaction	Refers to the overall feeling of goal accomplishment (i.e. does the student think that the instruction is satisfying and that is getting him closer to his goals?).

Table 3.1: Definitions of motivation model variables (traits and states)

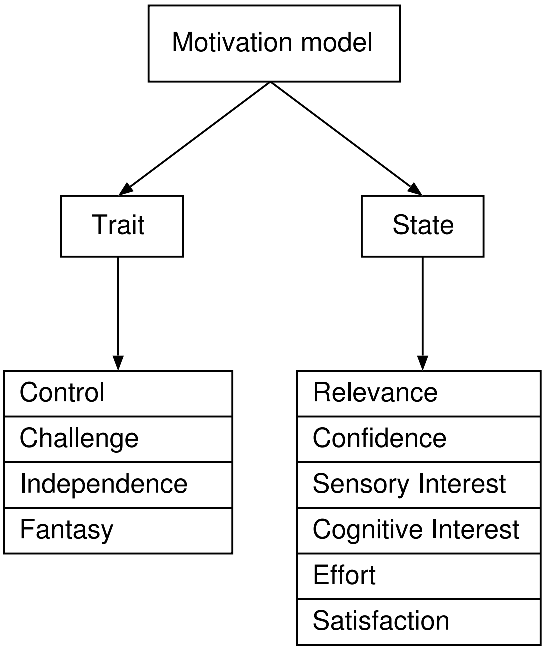


Figure 3.1: Motivation model

The trait variables aim to provide the system with a general picture of the goals it should pursue with a particular student. But to represent these personality characteristics as simple variables is, no doubt, a tradeoff between rigour and pragmatism. For example, a measure of how much fantasy a student likes during an instructional interaction is an oversimplification of all the complex aspects affecting this particular construct.

On the other hand, even a general and simple approach like this can help to create a better tutoring system, better 'tuned' to a particular student. Suggestions on how to deal with each of these variables have been sometimes obtained by performing studies with tutors teaching a particular subject to a particular group of students (e.g. mathematics to elementary-school children (Lepper *et al.*, 1993)). Whether these findings can be appropriate in other settings is an issue that the authors themselves raise: "Plainly, [. . .], the specific strategies that prove most effective with elementary-school children may not be the most effective for more advanced students. Young children, for example, may be especially susceptible to the use of fantasy or playful competition, whereas older students may find those techniques silly or dishonest" (Lepper *et al.*, 1993, p. 100). The provision of trait variables in our motivation model offers a simple solution to this problem, as students can let the system know about their major preferences for an instructional interaction.

The trait variables in our motivation model are: control, challenge, independence and fantasy. Theorists seem to agree on the importance of control and challenge for the student's motivation. Fantasy, although not often included in theories of motivation in Education, seems to be a factor that can play an important role in engaging the student (Malone and Lepper, 1987). Independence, as defined in table 3.1, is related to challenge, but also to the interpersonal motivations in Malone and Lepper's (1987) taxonomy: cooperation, competition and recognition. This concept was used by del Soldato (1994) in her motivation model, but here we place it as a trait characteristic that can guide a system using this motivation model to decide whether and how often it should offer help to the student.

The state variables all come from important factors recognised by theories of motivation, and they represent transient characteristics of the student that relate to the material being learnt. In figure 3.1 the state variables are presented in a more or less 'chronological' order. Thus, considerations of how relevant the task is to the student will likely happen before engaging in the task. This, together with how confident he feels about succeeding in the task, and the interest (both sensory and cognitive) that the lesson arouses in him, will influence the effort that he will put into the task. Satisfaction, as defined in table 3.1, represents the overall feeling of goal accomplishment, and will be influenced by all the variables above, plus the outcomes (as expressed by Keller's (1983) theory of motivation, summarised in section 2.4.1) of the task.

3.3 *How* to diagnose those motivational factors?

How can we get the information necessary to give values to the variables in our motivation model? Based on the work reviewed in sections 2.5.2 and 2.5.3, we focus on four different techniques:

1. *Questionnaires*, to gather information about the student's trait characteristics.

2. *Self-report*,

3. *Knowledge-based rules* (based on empirical studies), and

4. *Verbal communication*, to gather information about the student's motivational state throughout the interaction.

Questionnaire. Information about trait characteristics can be gathered easily through a simple questionnaire at the beginning of the interaction, although we should be aware of the difficulties of obtaining accurate data of people's attitudes (see, for example, Oppenheim, 1992).

Self-report. During the interaction, the approach of asking the student whether he wants help, or whether he is bored, etc. seems very attractive. It is probably one of the easiest methods to implement, and at the same time it is sometimes claimed that users can give subjective ratings easier than other explanations (Briggs *et al.*, 1996). On the other hand, as mentioned above, social studies research warns us of the difficulties of obtaining valid data regarding peoples' attitudes (Oppenheim, 1992). And even if we were able to obtain valid and reliable data through self-report, the question remains of how much such a method would interfere with instruction, and what the users' reaction would be to it. In order to study students' reaction to, and the usefulness of the self-report approach, we performed an empirical study in which a simple questionnaire and self-report was used as a way of providing the computer with information regarding the student's motivation. This study is discussed in detail in chapter 4.

Empirically based knowledge-based rules. del Soldato (1994) formalised and implemented a number of knowledge-based motivation diagnosis rules, but the rules used seem to have been based partly in theories of motivation in education and partly in common sense. As argued earlier, we believe that theories of motivation in education are not specific enough to inform the development of a motivation diagnosis component in an ITS. Therefore, in chapter 5 we focus on the formalisation of motivation diagnosis knowledge via an empirical study in which participants with tutoring experience were able to see replayed previous interactions of students with a tutoring system and were asked to predict students' likely motivational state. This, in turn, was translated into formalised knowledge-based rules to diagnose the student's motivational state.

Verbal communication. Language is a powerful communicator of affect, and it can be used in tutoring systems to empathise with students and to detect their emotional state during the interaction (e.g. (Allport, 1992; Horvitz and Paek, 1999; Person *et al.*, 1999; de Rosis and Grasso, 1999)). Creating a computational model of affective educational dialogue is out of the scope of this dissertation, but in chapter 7 we give some examples of how language could be used to detect the motivational state of an student.

3.4 Putting it all together

The outline for an Affective Tutor given above does not make use of all possible motivational factors or all possible techniques for their detection. Many other factors from other accounts of motivation (e.g. Csikszentmihalyi (1975); Spitzer (1995); Wlodkowski (1998)) and from previous motivation models implemented in ITSs (e.g. Matsubara and Nagamachi (1996)) could have been part of our motivation model. But we believe that our model presents a useful set of the main important characteristics for student motivation, while at the same time doing it with a small number of variables which makes it feasible to use.

In the following chapters we describe the empirical studies we performed to inform the development of the motivation diagnosis component of an Affective Tutor, focusing mainly on the self-report and the knowledge-based rules techniques. In order to evaluate this, we developed an Affective Tutor simulation, which is described in chapters 7 and 8. Obviously, an Affective Tutor needs not only a motivation diagnosis component, but also a motivational planning one. Since the core aspect of this dissertation is on the former, we based the motivational planning component of our prototype on the work of del Soldato (1994), as we explain in chapter 7. The evaluation of the motivation diagnosis component (by comparing the motivation diagnosis made by our prototype to that made by human teachers) is then discussed in chapter 9.

Chapter 4

Self-report study

4.1 Goal of study

As explained in section 3.3, the motivation for the study presented in this section was to study students' reaction to and usefulness of the self-report approach to motivation diagnosis. Thus, we developed a prototype ITS with an added facility that let students inform the system about their motivational state. By having a group of students interact with the system we set out to study:

1. What was the acceptance of the self-report method?
2. How often and when was the motivational model updated?
3. Which values did the motivational model variables take?

4.2 Materials

As mentioned above, in order to perform this study we developed a prototype ITS to teach Japanese numbers, with an added facility that lets students inform the system about their motivational state, which is described in detail in section A.1, appendix A.

The main interface of this prototype is presented in figure 4.1. With this study we wanted to find out whether having the option of reporting the self-perceived motivational state is intrusive for the students and whether this technique could be useful as a way to detect student's motivational state in ITSs. We were not aiming for engaging instruction with this study, rather, we wanted to know if self-report would be a viable option for motivation diagnosis. We aimed at offering to participants a variety of instructional situations that would hopefully result into clearly differentiable motivational states.

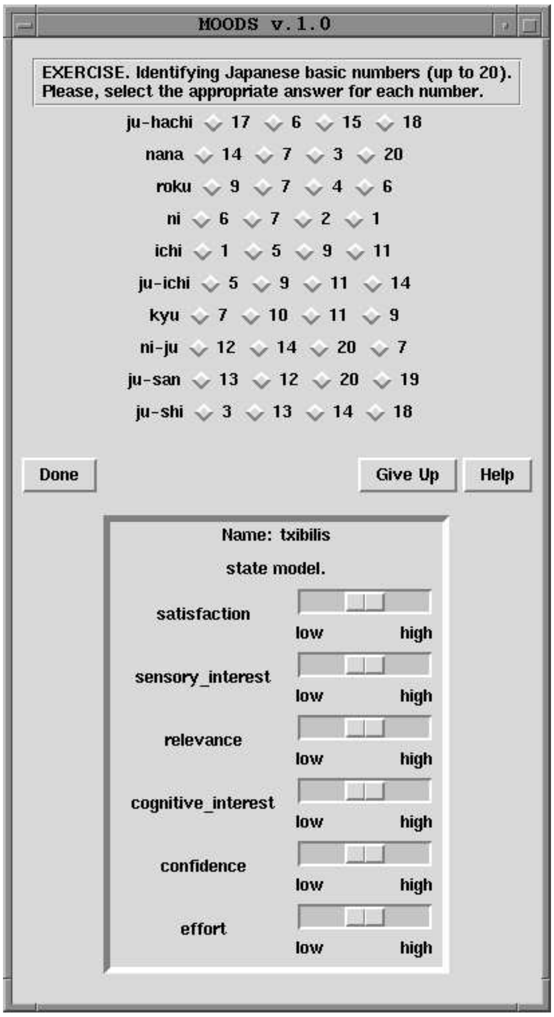


Figure 4.1: Main interface of prototype ITS system

With this in mind we organized the instructional units in the prototype ITS into six instructional paths, a description of which is given in table 4.1. A detailed description of the contents of each lesson is given in table 4.2. The instructional paths are

Description	Lessons order
1. Dry and difficult. First all the theory is presented, followed by the difficult exercises in which the student is asked to write the numbers in Japanese, and followed by the lessons to identify Japanese numbers.	Theory20, Theory100, Writing20, Writing100, Identi20, Identi100
2. Difficult. Similar to instructional path 1, but this time the student is given first the theory for the small numbers followed by the difficult exercise for small numbers. This is followed by a similar session about the numbers up to 100.	Theory20, Writing20, Theory100, Writing100
3. Difficult practice. Similar to instructional path 1, but this time the student performs the simpler exercises of identifying numbers before going to the difficult exercises.	Theory20, Theory100, Identi20, Identi100, Writing20, Writing100
4. Easy practice. Similar to instructional path 2 (first lessons dealing with numbers up to 20, and then lessons for number up to 100) but this time between the theory lesson and the difficult exercise, the student has to perform the easier exercise of identifying numbers.	Theory20, Identi20, Writing20, Theory100, Identi100, Writing100
5. More practice. Similar to instructional path 4, but extra practice exercises of identifying time and prices are given to the student.	Theory20, Identi20, Context20, Writing20, Theory100, Identi100, Context100, Writing100
6. Game with small numbers. Similar to instructional path 5, but instead of an easy extra exercise, the student can practice the acquired knowledge via a game before moving to the difficult exercise.	Theory20, Identi20, Game20, Writing20

Table 4.1: Instructional paths

Type		Lesson	Description
1.	Theory	Theory20	Explanation of Japanese number system (up to 20)
		Theory100	Explanation of Japanese number system (up to 100)
2.	Identification	Identi20	Identifying Japanese numbers (up to 20)
		Identi100	Identifying Japanese numbers (up to 100)
3.	Identification	Context20	Identifying time (up to 20)
	in context	Context100	Identifying prices (up to 100)
4.	Writing	Writing20	Writing Japanese numbers (up to 20)
		Writing100	Writing Japanese numbers (up to 100)
5.	Game	Game20	Game with numbers (up to 20)

Table 4.2: Description of instructional units

organized along a motivational continuum, from a very demotivating path number 1, to a pedagogically more appropriate approach, and therefore a more motivating one, in path number 6.

As we can see in table 4.1, the first instructional path presents first all the theory lessons (i.e. lessons *Theory20* and *Theory100* to learn both Japanese numbers up to 20 and up to 100), then the most difficult exercises in the instructional domain (i.e. lessons *Writing20* and *Writing100* to write the Japanese numbers up to 20 and up to 100), and then the easier exercises (i.e. lessons *Identi20* and *Identi100* to identify Japanese numbers up to 20 and up to 100, given the corresponding numeral).We expected that this instructional path would be very demotivating for the participants and would imply a decrease in the self-report of all the motivational factors (i.e. satisfaction, sensory interest, cognitive interest, etc.) as the interaction took place.

The second instructional path presents first the theory lesson *Theory20*, followed by the writing lesson *Writing20*. After this, the theory lesson *Theory100* is presented and then the writing lesson *Writing100*. This instructional path represents a pedagogically better approach, as the theory for the high numbers is taught only after the student had some practice writing the smaller numbers. Therefore, we expected that this instructional path would be slightly more motivating than the first instructional path.

Similarly, the remaining instructional paths follow this progression, and on the other side of the continuum we find the sixth instructional path which deals only with Japanese numbers up to 20, with a pedagogically more appropriate approach: first presenting the theory, then an easy exercise, then a game to consolidate the knowledge learnt and lastly the difficult exercise. We expected that this instructional path would create interest, pose a challenge, and through the game create a 'safe' environment (Spitzer, 1996) in which the students could practice their knowledge and which hopefully would increase students' overall motivation.

4.3 Methodology

In order to obtain participants for this study, we asked for collaboration to first-year university students at our Institution. They were told that their collaboration would help us in our research into tutoring systems, and that they would have to interact with a prototype tutoring system to learn Japanese numbers, although they were not required to know any Japanese at all.

As a result of our request, 18 first-year university students with no prior Japanese knowledge volunteered to participate in our study. Conveniently, having developed six different instructional paths, we allocated 3 students to each of the instructional paths. Although allocating only 3 students per instructional path meant that it would be impossible to apply statistical comparisons between results for each of the instructional paths, it was decided that it was more important to have participants in each of the instructional paths. The main point of the self-report study was to find whether self-report would be a viable option for motivation diagnosis, and it was crucial to test this against a variety of instructional situations, i.e. when the instruction was very engaging, when it was not engaging at all, etc.

On arrival at the room where the study was being performed, the participants were given three pieces of paper: one containing the instructions for the study, another with a screenshot of the prototype ITS interface and a third one with the definitions of the categories of the motivation model (these are reproduced in appendix A). After reading these and making sure they did not need any further clarification, the students were asked to fill in a small on-line questionnaire regarding the trait characteristics of the motivation model presented in section 3.2. This questionnaire is based on the definitions presented in table 3.1 and was developed by ourselves as a simple tool to measure the trait characteristics of the participants. The questionnaire can be seen in figure A.3. Once this questionnaire was filled in, the actual interaction with the system started.

Student interaction with the prototype ITS lasted for a varying amount of time for each student, ranging from only 8 mins 27 sec to 27 minutes (on average, interactions lasted about 14.5 minutes). After the interaction finished, the students were asked to fill in a questionnaire with 13 questions (given in section A.3), regarding mainly their opinions on the system and the usage of the self-report motivational model. For each question they could answer within a scale from 1 to 5 (1 meaning 'strongly agree' and 5 'strongly disagree'), and they were also encouraged to give extra comments. After they had filled in the questionnaire they were informally interviewed to find whether they wanted to provide any other comments regarding their interaction with the system.

As part of the instructions for the study, it was explained to the participants that the system would have a set of sliders representing various motivational factors, and they were encouraged to "use these sliders **as often as possible** whenever you think there is a change in any of these factors, since it is necessary for the computer to understand your current situation in order to modify the instruction accordingly." (emphasis in original instructions). This sentence was ambiguous on purpose in order to make the students believe that the computer would react according to their motivational state, as reflected in the settings of the motivational sliders. It was only after they had filled in the post-questionnaire and had offered any extra comments that they were debriefed on the true purpose of the study.

In total, the collaboration for each participant took about 30-40 minutes, and allowed us to gather a considerable amount of data, which was recorded thanks to the software TkReplay (Crowley, 1996). The results of the analysis of these data are presented in the following section.

4.4 Results

In order to study the usefulness of the self-report method, we set out to analyse the following issues, which are discussed in the following sections:

1. What was the acceptance of the self-report method as given by the answers to the participants' post-questionnaires?

2. How often and when were the different sliders updated?

3. Which values did the sliders take?

4.4.1 On acceptance of the self-report method

The first thing to study from the data collected for this study was the level of acceptance to the self-report method, as this could determine its usefulness in a real ITS. From the post-questionnaire given to the students, two questions relate directly to students' willingness to use the self-report method. Question 9 refers to the trait questionnaire and reads "I would prefer not to answer the trait questionnaire, even if it makes the instruction more efficient and personalised." Question 13 refers to the motivational state sliders and reads "I would prefer not to have to update the motivational state sliders, even if it makes the instruction more efficient and personalised.".

Students' answers to these questions seem to indicate that self-report could be an acceptable method for motivation diagnosis. The average answer to the question regarding the motivational sliders was 4 (where 1 means 'strongly agree' and 5 'strongly disagree'). As expected, the acceptance of the trait questionnaire (average answer was 4.21) was greater than that of the motivational factors sliders. This is probably due to the fact that the trait questionnaire is given only once at the beginning of the interaction as opposed to the sliders, which have to be updated throughout the interaction with the system.

Despite this, the acceptance of the use of sliders as a self-report method seems high given our expectations, but apparently students were not discouraged by having to update these sliders regularly. Nonetheless, this acceptance may be artificially high due to the length of the interaction with the system, which, as previously mentioned, was of only about 14.5 minutes on average. For instance, some of the students commented that longer use of the system would make the use of the motivational sliders less appealing. However, we believe that how the system reacts to the different values of the motivational sliders would also have a great impact in students' decision to continue updating the sliders or not.

4.4.2 How often and when were the sliders updated?

How often were the sliders updated?

In order to optimise the design of future self-report interfaces, it is important to understand which sliders were used more often. In table 4.3 we sort the different motivational factors according to the usage of their corresponding interface sliders. The column *Uncorrected total* shows the actual number of updates for each of the sliders. But by looking at the data, we find that some of these updates can be considered 'invalid', as they are immediately followed by an update of the same slider. This obviously indicates students' hesitation to settle on a particular value, and on these situations we would like to consider for our analysis only the last value chosen by the students.

	Uncorrected total	Total	%	μ	σ
Confidence	74	63	24.92	3.50	1.92
Effort	66	55	22.22	3.06	2.18
Satisfaction	52	46	17.50	2.56	1.85
Sensory interest	42	38	14.14	2.11	1.28
Cognitive interest	35	33	11.78	1.83	1.25
Relevance	28	24	9.43	1.33	1.37
Total	**297**	**259**	**100.00**	**14.39**	**7.08**

Table 4.3: Slider usage

Thus, in order to calculate the corrected values (column headed as 'Total'), we did not take into account the sliders updates which were followed by an update of the same slider in a lapse of time equal or less to 2 seconds. For all subsequent analysis, it is these corrected data that are used.

As we can see, the *confidence* motivational factor is the one that was updated more often (amounting to nearly 25% of all updates). This was expected, and it is in accordance with the suggestion made by Briggs *et al.* (1996) that self-confidence is a factor easy to report. On the other side of the scale we see that *relevance* was updated very rarely (on average only 1.33 updates per student). The updates for the rest of the motivational factors were between these two extremes.

This seems consistent with the answers to question 11 of the post-questionnaire, which deal with the ‘ease’ of updating the motivational sliders. Question 11 reads “The sliders representing motivational state factors were easy to answer (I easily know my ‘level’ for each of these factors)” and the average answer was 3.05 (again, 1 meaning ‘strongly agree’ and 5 ‘strongly disagree’). However, participants were also asked to comment on whether they found that any of the motivational factors was particularly difficult to answer. Only four people answered this, *relevance* being particularly difficult to answer for all four, and *confidence* and *cognitive interest* by one of them. The average answer to question 11 without taking into account those students who found *relevance* particularly difficult to answer drops to 2.21. Thus, it would seem that except for the relevance factor, students did not have great difficulty updating the motivational slides.

Does acceptance of the self-report method affect slider update? Although most of the participants would be willing to use self-report facilities if this would make the instruction more efficient (as seen in section 4.4.1), there remains to be seen whether students’ acceptance of the self-report method has any influence on the number of updates to the motivational sliders that they make.

In figure 4.2 we can see the relation between the acceptance of the self-report method and the number of slider updates made by the participants. As it can be readily appreciated, the variety of number of updates is very great, and there is no clear relation with the level of self-report acceptance. Actually, if we consider the average number of updates for each of the possible levels of acceptance, there seems to be, somehow surprisingly, a tendency to update less often the motivational sliders as the

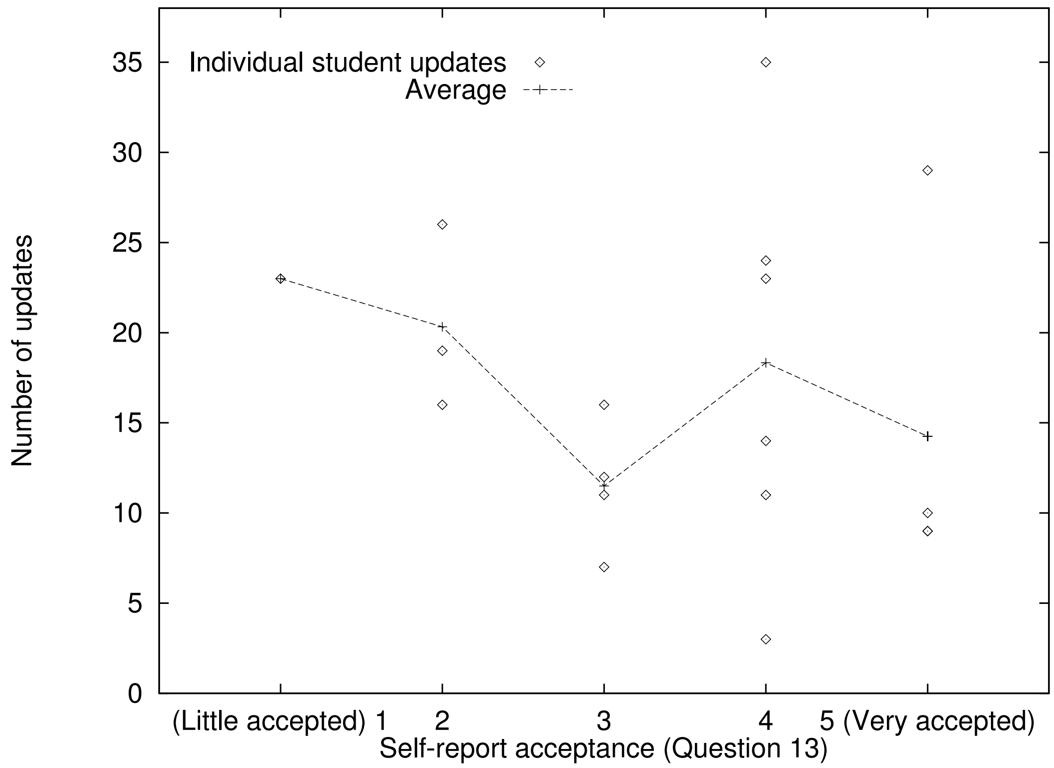


Figure 4.2: Relation between self-report acceptance and the number of slider updates

self-report acceptance increases. Nevertheless, the diversity of the amount of updates is too great and it seems clear that the level of self-report acceptance alone cannot explain this diversity.

When were the sliders updated?

As mentioned earlier, in the version of the prototype used for this study, the motivational sliders were available at all times during the interaction. But when were they used by the participants? This is an important issue to consider, as the self-report interface could be perhaps more efficient if the motivational factors were only available to the students when they are more likely to use them.

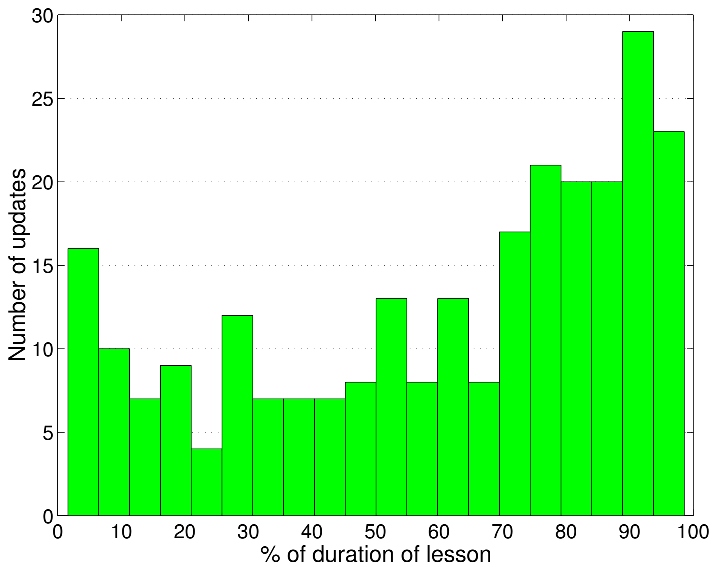


Figure 4.3: Distribution of slider update totals through duration of lessons.

In order to analyse this we plotted all the occurrences of slider updates against a normalised duration of a lesson. The result of this can be seen in figure 4.3. There we can see that a total of 16 slider updates were made during the initial 5% of a lesson. This figure does not separate between different types of motivational sliders or lessons, but it provides a good indication of when slider updates occurred during the lessons. As another example, we can also see that the time when most sliders updates took place was nearly at the end of the lessons, namely during the 90-95% time percentage window.

Figure 4.4 represents the same type of histograms, but separating the update occurrences by sliders. Thus, figure 4.4(a) represents the distribution in time of all the 'Confidence' slider updates. There we can see, for example, that the smallest frequency of updates of the confidence motivational slider happened during the 30-60% time percentage window of the lessons.

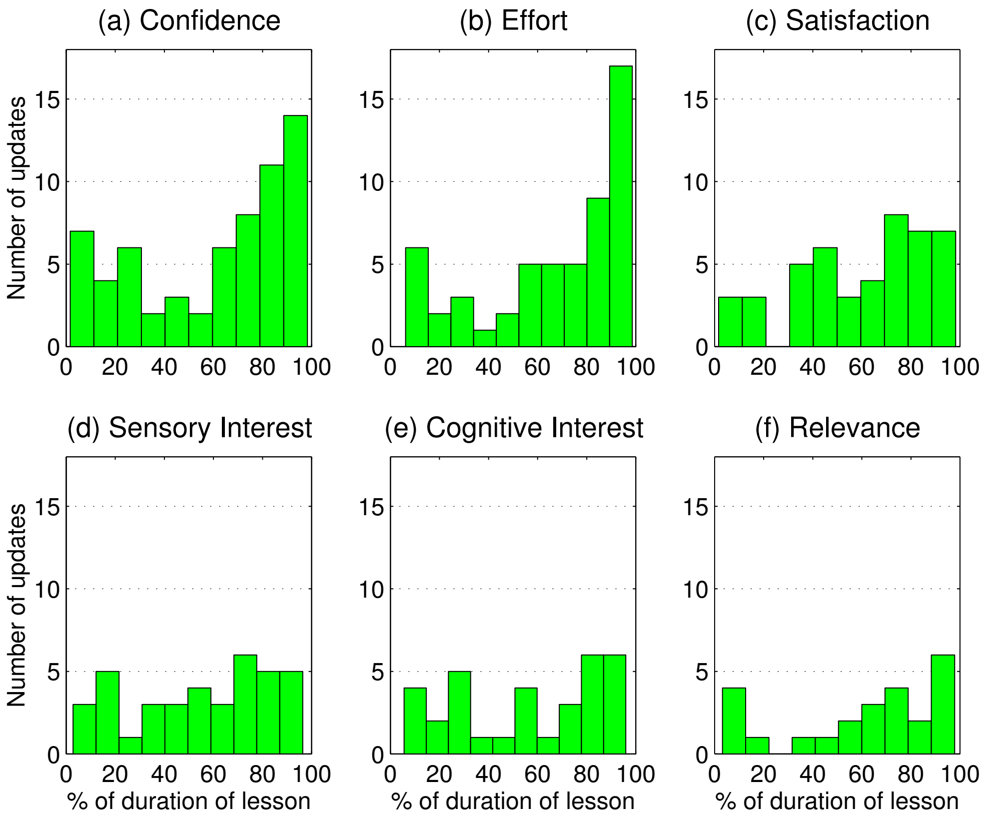


Figure 4.4: Distribution of individual sliders updates through duration of lessons.

A number of interesting conclusions (which are summarised numerically in table 4.4) can be drawn from these figures:

- The biggest number of motivational slider updates happened during the last part of the lessons. In fact, 49.42% of all updates were made during the 70-100%

	Main		Secondary		Rest
	Value	Window	Value	Window	Value
Effort	47.27%	80-100%			52.73%
Confidence	58.73%	65-100%	26.98%	0-30%	14.29%
Satisfaction	47.83%	70-100%	30.43%	30-55%	21.74%
Sensory Interest	42.00%	70-100%	21.05%	0-20%	36.95%
Cognitive Interest	33.33%	5-30%	24.24%	80-100%	42.43%
Relevance	50.00%	70-100%	16.67%	0-10%	33.33%

Table 4.4: Main times of motivation sliders updates

duration of lessons. This would point towards a self-report interface that should be made available towards the end of each lessons, although the analysis of figure 4.4 suggests a more elaborate picture.

- **Effort**. The updates to this slider clearly happened mostly at the end of the lessons. A detailed look at the data shows that 47.27% of the updates were made during the 80-100% duration of lessons.

- **Confidence**. In the case of the 'Confidence' slider, the largest frequency of updates is also towards the end of the lesson (58.73% of updates in the 65-100% duration window), although there is also a substantial amount of updates at the beginning of the lessons (26.98% of updates in the 0-30% duration window). This can be interpreted as being a motivational factor that was understood by participants in two possible ways:

    1. What is your confidence in performing this lesson correctly?
    2. How confident are you that you *performed* this lesson correctly?

- **Satisfaction**. As can be seen from figure 4.4(c), the satisfaction slider was updated mainly around the middle and the end of the lessons (30.43% of updates in 30-55% duration window and 47.83% of updates in 70-100% window, respectively), but the difference in frequency is not as high as in the data for 'Confidence' and 'Effort'. Actually, we can interpret that the satisfaction slider was mainly updated during the second half of the lessons (73.91% of updates in 45-100% duration window).

This is consistent with the intended meaning for this motivational factor, as the satisfaction could vary throughout a lesson (for example, if the initial confidence of the student proves to be overstated as he realises the actual difficulty of the lesson).

- **Sensory Interest**, **Cognitive Interest**, and **Relevance**. These three factors have fewer updates than the other motivational sliders and therefore it is more difficult to draw generalisations on their data. Nevertheless the three of them seem to have a similar pattern of usage, in the sense that most of the updates happen around the beginning and end of the lessons. (21.05% of updates in the 0-20% window and 42% of updates in the 70-100% window for 'Sensory Interest'; 33.33% of updates in the 5-30% window and 24.24% of updates in the 80-100% window for 'Cognitive Interest'; 16.67% of updates in the 0-10% window and 50% of updates in the 70-100% window for 'Relevance'). Nevertheless, 'Sensory Interest' and 'Cognitive Interest' have also a substantial number of updates around the middle of the lessons. Given this variety, and the small number of updates for these factors, it looks like a viable option for self-report interfaces would be to include these factors as options that the student could update throughout the interaction, but that are not present in the main interface of the Affective Tutor. We further discuss this issue in section 4.5.

4.4.3 Which values did the sliders take?

The values that the sliders took during the interactions with the participants can provide us information about two important issues, which we cover in the following sections:

1. Is the scale used for the different motivational factors appropriate or should it be modified?

2. Can we trust the information about motivation that the students provide?

Is the scale used to represent the different motivational factors appropriate?

In order to study the appropriateness of the scale used, we can look at the distribution of updates along the different values of the scale given. This can be seen in figure 4.5. In it we represent the five possible values that the sliders could take by the numerical values used by the prototype ITS (-10 for the lowest, and 10 for the highest value), and for each of them we draw a bar representing the number of times that the slider was moved to that position. For example, in figure 4.5(a) we see that the Confidence slider was moved to position 10 (the highest position in the scale) a total number of 5 times. And in figure 4.5(f) we can see that the Relevance slider was never moved to the lowest point in the scale (0 number of updates for the -10 scale value).

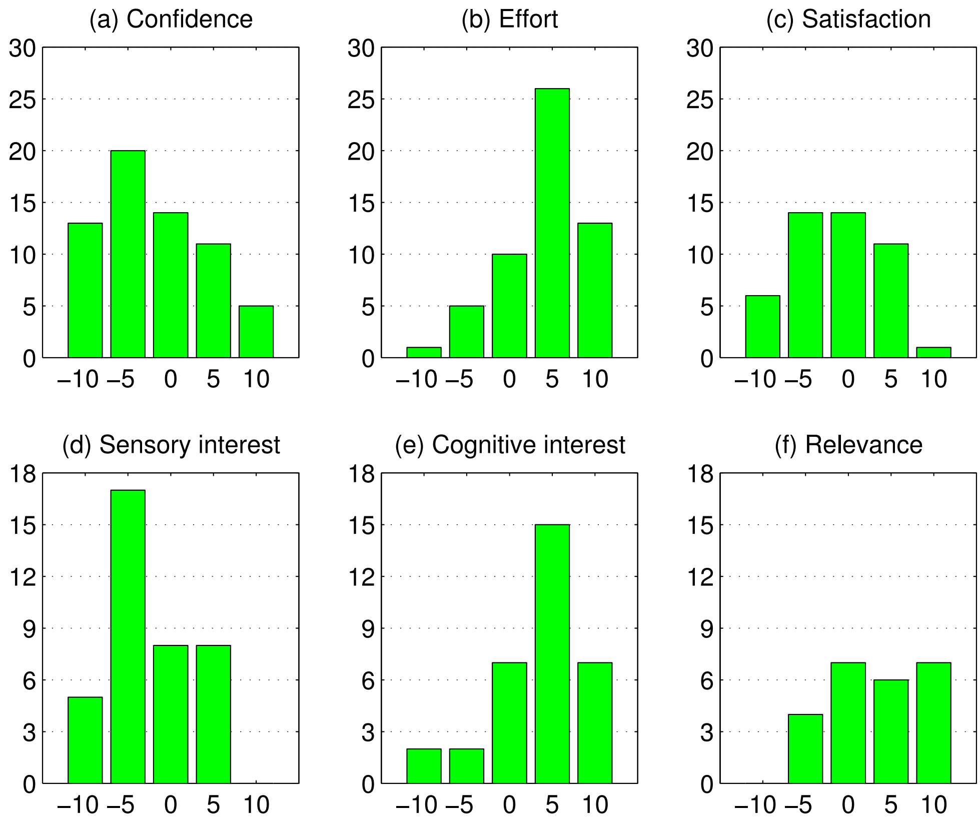


Figure 4.5: Total number of updates per value of slider scale

From figure 4.5 we can see:

1. that virtually all the possible values in the sliders' scale were used;

2. more importantly, that the distribution of all of them (except for Relevance) resembles an inverted 'U' shape. If the extreme values of the slider's scale were chosen very often by the students it could indicate that the scale would require more points in that direction. But in our case we can see that the peak of the scale value distribution is in all figures around the central or one of the adjacent points in the scale.

Thus, these two points seem to indicate that the scale used for this study is appropriate for our purposes, and that we should not subtract or add any more points to it.

Can we trust the information obtained through the motivational factors sliders?

When asking people about their emotions, orientations, etc. there is always the danger that the answers might be false, or at least inaccurate (see Oppenheim, 1992; Reeves and Nass, 1998). In our study, how can we find out whether their interaction with the motivational sliders was truthful and that the data so collected really represents their motivational state? It is impossible to prove this, but we can approach this issue in the following way. The participants were presented with certain lessons organised in a number of instructional paths. By applying our pedagogical knowledge as teachers and the motivational theories presented in chapter 2, we can predict what the motivational state of the participants 'should' be at certain points during the interaction with the prototype ITS. If none of these predictions match the data obtained through the self-report interface, we can assume that either our predictions or the data given by the students is not accurate. If some of the predictions are proved right by the data obtained, then we will have some assurances that the self-report interface provides, at least sometimes, useful information about the student motivational state.

As explained in section 4.2, the six instructional paths were designed attempting to create a 'motivational continuum', starting with a very demotivating one (number 1) and ending with a much more appealing one (number 6). We can check if this was the case by looking at the average values of the motivational variables throughout all the lessons for each path. Looking only at the exit values of each motivational variable (i.e. the value at the end of the final lesson of each path) would only give us a very limited view of the student's motivation for each path. For example, if the last lesson in the instructional path number 6 were very satisfying, but the previous lessons were not, we could obtain a very high "satisfaction" exit value, which would not give us an accurate picture of the student's satisfaction during the instruction. On the other hand, looking at the average values of the motivational variables throughout all the lessons in each instructional path, gives us a better indication of the overall motivation for each path.

Thus, we present in figure 4.6 the average value of four of the motivational factor sliders according to the different instructional paths. Although the five possible values of each motivational slider are presented to the participant as values in a scale between 'low' and 'high', the program stores these values as numbers in a scale between -10 and 10 in steps of 5 (as already seen in figure 4.5). It is these values that are used in order to calculate the averages.

The *relevance* data is not included in figure 4.6, as this is a factor that applies more to individual lessons than to a whole instructional path, and in any case, as we mentioned earlier, the *relevance* slider seemed difficult to update and it was the slider with the lowest number of updates. Similarly, we do not include *effort* in the figure, as it is difficult to predict what its values 'should' be. It is reasonable to expect that as the instruction becomes more motivating the students would be willing to put more effort into the task. But at the same time, some of the lessons in some of the 'demotivating' instructional paths are very difficult, which would require a very high effort on part of the participants.

From the four factors pictured in figure 4.6, our expectation seems to be confirmed mainly for *cognitive interest* and *sensory interest*. The other two factors (*satisfaction* and *confidence*) show a timid increase up to Path 5 and a sharp decrease in Path 6. The pattern for *satisfaction* is slightly puzzling, since the participants commented that they quite enjoyed the game, although we should note that except for the value of Path 5, Path 6 shows the highest 'Satisfaction' average of all the paths. Perhaps these data were influenced by the fact that the game in Path 6 was actually quite engaging and

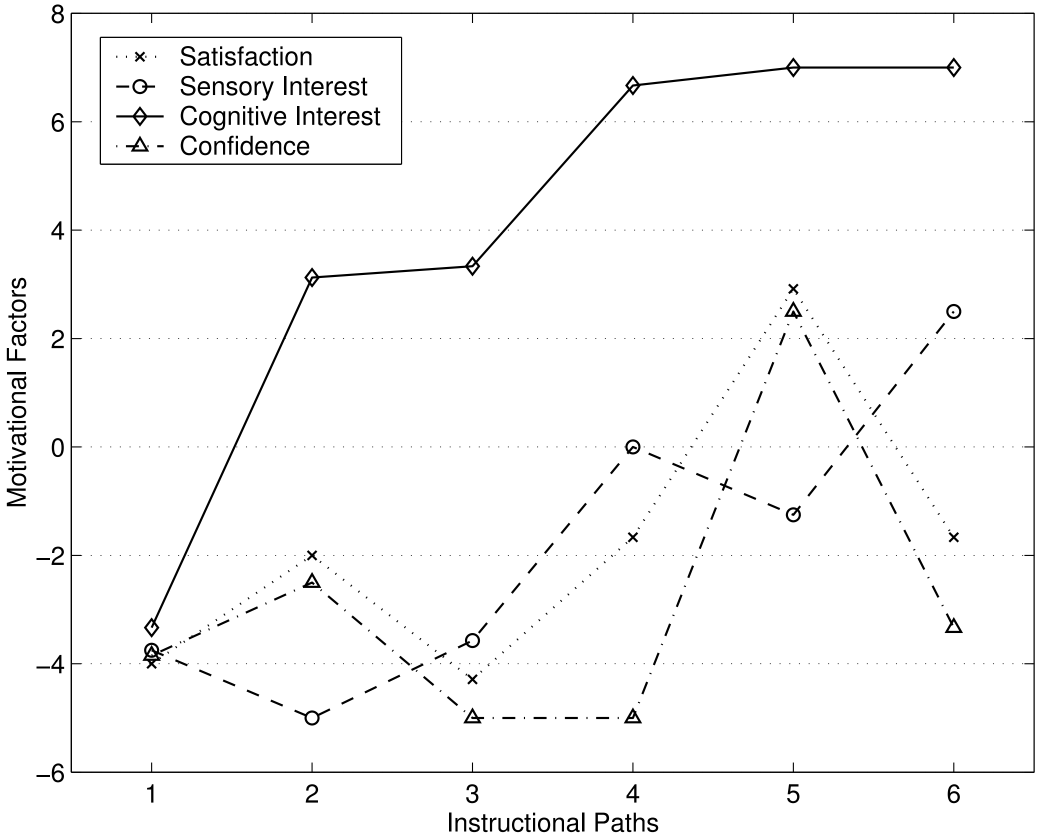


Figure 4.6: Sliders' average values for the six instructional paths

participants did not concern themselves much with updating the slider for these factors.

Also apparent in figure 4.6 is the positive relation between *confidence* and *satisfaction* (both of them very low, on average below 0 except for Path 5). Apparently, the curriculum taught by the prototype ITS was quite difficult, as reflected by participants' *confidence*, and this may explain that positive relation. Lack of confidence in solving a task can bring anxiety, and a lack of motivation in general. On the other hand, a very high degree of confidence can bring boredom and also lack of motivation in general. But our data only provides (on average) low levels of confidence so we cannot check this against our data. It would have been interesting to have participants with a high level of Japanese knowledge (and therefore an extremely high level of confidence for the tasks given in this study) in order to check this hypothesis.

More interesting still is to check if there are variations in the motivational state of the participants for individual lessons according to the path that they belong to. For example, the lesson "Writing20"[1] appears in Instructional Paths 2 and 5 (among others). In the Path 2 (see details of Instructional Paths in table 4.1) this lesson comes immediately after the first theory lesson on Japanese numbers. On the contrary, in Path 5 the student is given two easier practice lessons before coming to lesson "Writing20". Therefore, we would expect the confidence level in the Path 5 to be higher than in Path 2.

By following a similar procedure, we can make a number of predictions about the likely values of certain motivational factors during a particular lesson, depending on the instructional path where they occur. We present three of these predictions below:

a) The lesson Theory100, in which theory about the Japanese numbers up to 100 was explained, appears in all but the last instructional path. In paths 1 to 3 this lesson is presented immediately after the theory for small numbers. On the contrary, in path 4 the participant performs two exercises on small numbers before moving to this lesson. And in path 5 he has three exercise lessons prior to this one. Therefore we would predict that his confidence for this lesson would be higher in paths 4 and 5.

b) Similarly, the lesson Identi100 in path 5 is performed after the participant had the chance to practice with smaller numbers more than in the other paths and in a more coherent pedagogical style, so we would expect both his confidence and satisfaction to be higher.

c) The third prediction involves the lesson Writing20, which, as for the other two examples, is presented in path 1 right after the theory lessons. On the contrary, the student has more practice with easier exercises as the instructional path number increases. Therefore we can predict that his satisfaction would be higher for the high numbered instructional paths.

These predictions were largely met, as can be seen in figure 4.7, where we present

[1] In this lesson the student is asked to write the Japanese number in text, given the number in figures.

the actual data obtained during the self-report study that relates to the three predictions stated above:

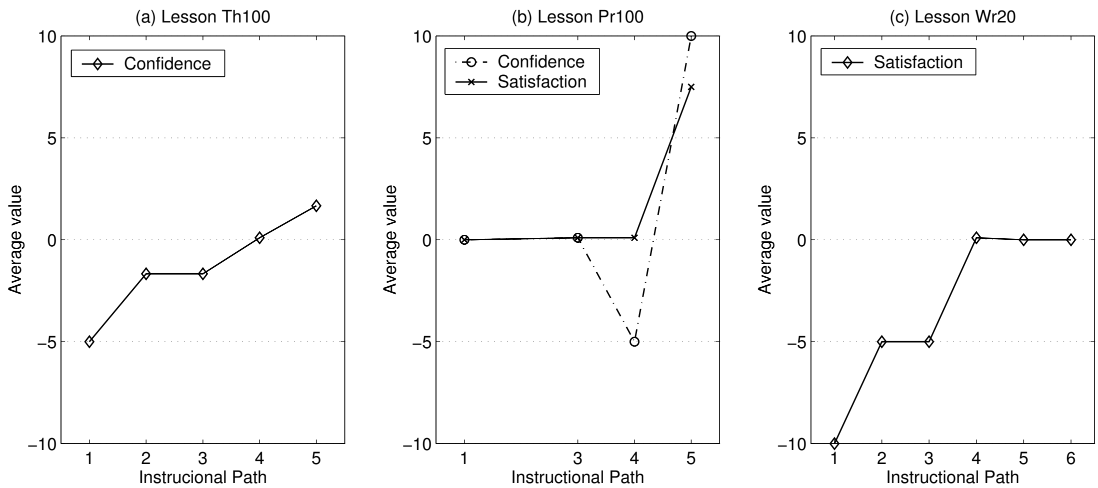


Figure 4.7: The effect of instructional paths on participants' motivational state .

4.7(a) In this figure we can see that prediction (a) was clearly met in the self-report study. As predicted, participants' average confidence for the lesson Theory100 would be higher in paths 4 and 5, and this clearly appears to be the case, as can be seen in figure 4.7(a).

4.7(b) In this figure, we can see that, as stated in prediction (b) above, the participants' average confidence and satisfaction for the lesson Identi100 in path 5 is higher than for other paths, although we should also note that we get an unexpected result, as the confidence for this lesson in path 4 seems to be lower than for paths 1 and 3.

4.7(c) In this figure we can see that prediction (c) was also met, as it is clear that the participants' average satisfaction for lesson Writing100 was higher for the high numbered instructional paths.

4.5 Discussion

This study offered an important insight into the issue of the usefulness of the self-report method for motivation diagnosis. The results obtained from this study give useful clues as to how this method could be used in a real ITS, and we discuss them in this section. The study results also give indications of what changes could be made to the self-report interface to make it more efficient, and we provide in section 4.5.1 a summary of these.

One of the main doubts about the self-report method before conducting this study, was that of its acceptance. How would students react to it? Would it be used? Would it provide useful information about students' motivational state? Given the results presented in section 4.4, we do believe that the method of self-report could be used satisfactorily for motivation diagnosis in ITSs. We have seen that participants of this study seemed to accept the use of the self-report interface and that its use provided (though not always) useful information about their motivational state.

Participants seemed to think that self-report could be a good method for communicating with the computer about motivational issues, and a method not intrusive on their learning. But also, as noted in the previous section, some of the students commented that a longer period of use of the system may make them lose interest in using the self-report facilities.

As we suggested, the reaction of the system to the values of the motivational sliders may have a great impact on the willingness to continue using the self-report facilities. Since the prototype ITS used did not react to the motivational sliders values and the interaction was quite short, we can only speculate that, although the method seems acceptable, care should be paid to making the reaction of the system to users updates of the sliders very obvious, in order to encourage the use of the self-report facilities.

In this study, for example, one of the questions posed to the student was: "The system seemed to react appropriately when I updated the motivational factor sliders". The average answer to this question was 3.44 (where 1 meant 'strongly agree' and 5 'strongly disagree'), making clear that the participants did not see much reaction from the system when they updated the sliders. This lack of system reaction may have affected participants' willingness to use the self-report facilities.

Students' perception of how appropriately the system reacted to his motivation model could be useful in a formal evaluation of a 'motivating' ITS. After all, as it is noted in the motivation literature, it is the feeling of control, rather than control itself, what seems to be important in motivating students. The perception that the system is reacting appropriately to one's self-reported motivational state may be regarded as a kind of indirect control.

In this sense, it is interesting to note that some students commented that they changed some sliders in an exaggerated way to try to make the computer react to their inputs. In these cases, a speedy and obvious reaction would be very important, as it would give the students the feeling that the system reacts appropriately to their actions.

At the same time, the results obtained in section 4.4 suggest some changes in the self-report interface, which would make it more efficient and therefore, more likely to be used by students. Several students commented how they mostly updated the sliders at the end of each exercise. Others commented how it was good to have them available at all times. For instance, the *confidence* could be high at the beginning of one exercise, but lower later during the same exercise, when realizing that the exercise was actually more difficult than expected. A compromise between these different approaches is outlined below.

First of all, we should try to avoid those factors that were poorly understood or barely used. In our study we have seen that "Relevance" was very seldom updated. Although providing relevant material is a major factor for creating motivating instruction, it seems clear from this study that to question students about each particular lesson's relevance does not seem to be appropriate and/or useful. With a limited curriculum like the one present in the prototype, all the lessons are relevant to the task of Japanese number learning, so the relevant motivational slider can be considered superfluous and it is probably best to consider the creation of relevant material simply as a curriculum design issue.

At the same time, we could use the infrequently used factors for occasional problem detection rather than for continuous update by the students. In this sense, 'sensory interest' and 'cognitive interest' should be available to the student in case he wants to update them, but not as an integral part of the self-report interface. They should be hidden from the main self-report interface, but accessible to the student via a menu in case he wants to update them. This would make the interface much simpler, but offering similar informative capabilities.

Therefore, the main self-report interface could be reduced to the three factors that are easier to update and more regularly updated in our study: ‘confidence, ‘effort’ and ‘satisfaction’. But at the same time, we should make these available at the times when they are most likely to be updated, and when their interpretation is clearer. Thus, ‘satisfaction’ should be available at all times, being the overall index for student’s motivation. ‘Effort’ should be available at the end of each lesson, and ‘Confidence’ at the beginning and end of each lesson, in order to be able to appreciate discrepancies between the initial expectations of the students towards the lesson, and the confidence in his actual performance at the end of the lesson. Providing these factors at these particular times would limit the expressiveness of the students, but it would give a clearer interpretation of the meaning of each motivational factor slider.

4.5.1 Suggested changes to the self-report interface

To summarise, the present study would suggest that the following changes could improve the self-report interface:

1. Only the slider ‘Satisfaction’ should be available at all times during the interaction with the system.

2. The system should ask explicitly about student’s confidence in performing a given task at the beginning of each lesson, and about his confidence in his performance at the end of the lesson.

3. At the end of the lesson he should be questioned about the effort he made in performing the given tasks.

4. The sliders corresponding to the motivational factors ‘Sensory Interest’ and ‘Cognitive Interest’ should not be present in the main self-report interface, and would not be queried explicitly, but the student should be able to access and update them at any time during the interaction if he so wishes.

5. The slider corresponding to the ‘Relevance’ factor would be better completely ignored for the self-report interface, as it proved to be confusing and not very informative.

Chapter 5

Motivation Diagnosis study

5.1 Introduction and goal of the study

The issue of how human teachers detect their students' motivation seems to be taken for granted, and has been virtually unexplored in AI-ED research. Introspection and observational studies could throw some light on this issue, but they may be of limited usefulness for motivation diagnosis in ITSs. In a 'traditional' instructional setting or any other social interaction there is an incredible amount of information available through various communication channels, such as facial cues, intonation, posture, etc. Davis (1976). Many cues that help us detect other people's motivation are perceived unconsciously via these channels, which makes it difficult to elicit motivation detection knowledge.

In order to limit the number of sources of information available for knowledge elicitation, we designed a study in which a participant will see exclusively the screen interaction of a student with an instructional system. That is, the participant will be able to see in a computer screen only the interface of the instructional system with which the student is interacting.

We wanted to be able to extract and formalise tutors' knowledge about motivation detection, and we expected that it would be easier for tutors to rationalise their motivation diagnosis knowledge in this setting than if they were presented, say, with video-recordings of tutoring interactions. At the same time, we believed that the knowledge thus inferred would also be easier to formalise in terms of the information available to the instructional system (such as time of interaction with system, mouse movements, etc.).

The approach followed in this and the following chapters in order to elicit the knowledge about motivation detection and the development of the Affective Tutor simulation can be compared to that used in a knowledge engineering process, which is usually made up of five steps (Turban, 1992):

1. Knowledge Acquisition, which involves the actual capture of the expert knowledge. This step is covered by the first part of this study, in which we recorded the comments made by the participants while watching the interaction of a student with an instructional system and created a first approximation of the rules representing the knowledge about motivation detection. A sample of the knowledge thus acquired is given in table 5.3.

2. Knowledge Representation, which involves the representation of the knowledge previously captured. This step is covered by the second part of this study, in which we developed a final set of motivation diagnosis rules based on the knowledge previously acquired. This set of rules is presented in tables 5.4 to 5.13.

3. Knowledge Validation, which involves testing the elicited knowledge. This step is covered by the Motivation Diagnosis Validity Study presented in chapter 6.

4. Inference, which involves the development of software that incorporates the elicited knowledge. This step is covered by the development of MOODS (a prototype affective tutor simulation) which is described in chapter 7, demonstrated in chapter 8 and evaluated in chapter 9.

5. Explanation and Justification, which involves the development of software that would allow answering questions connected to the way in which the computer uses the knowledge. This step is only partially covered by the development of MOODS, which can give basic explanations of its reasoning. In section 10.3 we suggest as further work a more sophisticated explanation mechanism.

5.2 Materials

In order to perform this study we used A_MOODS, an augmented mode of the prototype system used in the self-report study described in chapter 4. The augmented version used for this study (A_MOODS) is a version that allows us to replay the interaction of a student, while a participant can infer changes in the student's motivational state. This is explained in more detail in section 5.2.1.

5.2.1 A_MOODS description

The goal of this study was to explore issues of diagnosis of students' motivation during instructional interactions. Therefore, we wanted the participants to watch recorded interactions of a student with an ITS, and to infer and to comment on the affective state of the student during the instructional interaction.

In order to do this, we developed A_MOODS, which can be used to replay the actions of a previous student's interaction with the prototype ITS previously developed[1]. The A_MOODS interface can be seen in figure 5.1.

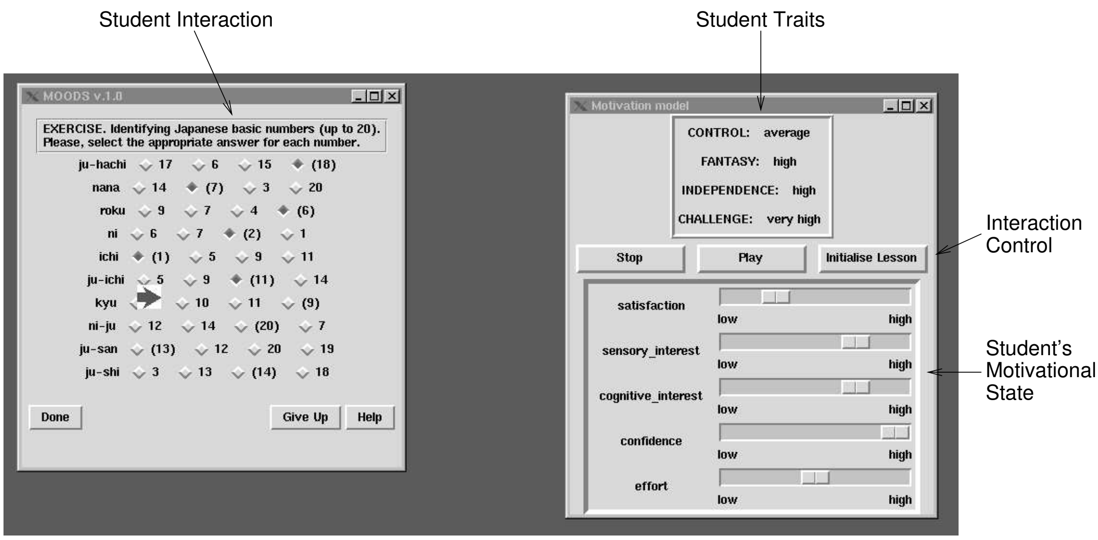


Figure 5.1: A_MOODS interface.

[1]The recording and replaying facilities were possible thanks to the program TkReplay (Crowley, 1996).

The window to the left (with title MOODS v.1.0) is where the actions made by the student are replayed, except his interaction with the motivational sliders. The interaction of the student with the motivational model was excluded from the A_MOODS interface to avoid interference with the diagnosis made by the participants of this study. We wanted to find how much information about a student's motivational state could be elicited by looking at his interaction with a tutoring system, but not necessarily one with self-report facilities. At the same time, the replaying of the student's interaction with the motivational model would have interfered and added confusion to the A_MOODS interface and to the task of the participants.

Due to technical difficulties with TkReplay, the actual mouse pointer used by the student could not be used during the replay. In its place, an arrow (around the centre of the window in figure 5.1) indicates the original mouse movements[2].

The window to the right (titled Motivation model) consists of three frames.

1. The top frame is a representation of a number of *student traits*, which provides information about some general learning characteristics of the student. These were obtained through the trait questionnaire of the self-report study (see chapter 4).

2. The middle frame contains three buttons in order to control how the student interaction is replayed.

[2]It is important to note that TkReplay does not record all the mouse movements. It simply records the mouse events in relation to the components (buttons, check boxes, sliders, etc.) of the interface. Thus, if the mouse is on top of a button and the user moves it to another button, TkReplay would record when the mouse exits the first button and when the mouse enters the second button, but the actual path followed by the mouse between these buttons is not recorded. When replaying the interaction, the path followed by the arrow between these buttons will be interpolated by the program to simulate the original movement.

This lack of precision in the recording of the mouse movements can be important if there is a large blank area between interface components, since the straight line followed by TkReplay and the actual movements made during the recording could vary substantially. But when there are no large blank areas between interface components, the importance of this lack of precision can be considered negligible. This is the case in the interface used for this study, since the blank space between different components is very small, and therefore the replay of the mouse movements approximates quite accurately the actual mouse movements made by the students.

- The replay of the student interaction starts by pressing “Play”and stops by pressing “Stop”.
- The button “Initialise Lesson” allows the participant to ‘rewind’ the interaction to the beginning of the current lesson.
- The replay of the interaction is done in real time, except for the replay of the theory lessons. In this type of lesson there is very little student activity, and therefore the system simply shows a message informing how long the student took to learn the lesson (see figure 5.2), and afterwards the interaction continues.

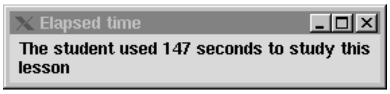


Figure 5.2: Sample message to inform the participant of time spent in theory lesson.

- In order to let the participant update the student’s motivational model, the interaction stops by default in the following three cases:
  (a) When the student presses any of the buttons (Done, Give Up, or Help), but before any feedback is given by the system.
  (b) When feedback is presented to the student.
  (c) When a new lesson is presented to the student.

Figure 5.3: Sample message to remind the participant to update the model.

In these cases a small message similar to that in figure 5.3 will be shown to remind the participant to update the motivational model. But if the participant did want to update the motivational model or comment on any aspect of the instruction at any other time, he could stop the interaction by pressing the “Stop” button.

3. The bottom frame is a representation of the student's *motivational state*, as presented in chapter 3. The task of the participants for this study was to predict the likely values of these motivational variables as the instruction replay took place.

5.3 Methodology

In order to obtain participants for this study, we asked for collaboration among post-graduate students at our Institution. Since the data obtained from this study was only preliminary and would be validated with a further study (see chapter 6), the only prerequisite for participation was to have at least one year teaching and/or tutoring experience. As a result of our petition, 10 post-graduate students volunteered to participate in the study. All the participants had high computing skills but had no Japanese knowledge (except one of them, who had studied some Japanese many years ago, although she had forgotten most of it), although this was not a problem, since A_MOODS was also modified in order to show the correct answers to the exercises. This can be seen in figure 5.1, where the correct answer to each question is between brackets. For example, the first question "ju-hachi" has the correct answer "18" between brackets. This helped the participants of our study to find if the students were giving the correct answers to the exercises.

Ideally, we would have wanted the collaboration of 12 participants, so that we could allocate two participants per instructional path. Since this was not possible, we allocated two participants per instructional path, except for paths 1 and 2, which only were observed by one participant each. For each participant, once the instructional path was selected, the actual recording to observe was selected randomly among those obtained as part of the self-report study corresponding to the given instructional path.

Each participant was told that his interaction with A_MOODS would last approximately 40 minutes. Since the interaction could be stopped at any moment that the participant wanted to give some comments, this meant that the number of lessons viewed by each participant would not be necessarily the same. This was also the case, since some student recordings did not convey as much information as others, and therefore some lessons would be quickly skipped by the participants, while other lessons would require a long time to give comments.

Before the interaction with A_MOODS took place, the participants were given an instruction booklet, which is reproduced in appendix B. After reading this, the actual interaction with A_MOODS started, which can be summarised as follows:

1. The participant was presented with information about certain trait characteristics of a student (who had no knowledge of Japanese before the interaction took place).

2. Then she was shown a replay of the student's interaction with the prototype ITS.

3. Throughout the interaction, and particularly at any stop points, the participant was encouraged to give verbal comments (which were recorded for analysis) on the student's motivational state and the possible factors affecting it.

4. Whenever the interaction was paused, the participant was asked to update the motivational state variables if she thought she had enough information to make an inference. At the same time, the participant was asked to verbalise the reasoning behind her inferences in as concrete terms as possible.

5. When the student pressed any of the three buttons available to him (Done, Give up, or Help) the participant was also encouraged to comment on the type of feedback that she thought would be the most appropriate to give to the student at that particular moment.

5.4 Results

In this section we present the results from the study. But before we proceed to the actual inferences about motivation diagnosis knowledge, it is interesting to note some general impressions about the study.

On the feasibility of inferring knowledge through the basic interface. It was a common comment among the participants before they started the study that the task was very difficult, and that it would not be possible to make any inferences based on the information provided. As we said earlier, in face-to-face teaching the tutor has an incredible amount of information at her disposal to infer the student's motivational state (his facial expressions, posture, etc.). But contrary to their own expectations, most of the participants made a considerable number of reasoned inferences about the students' motivational state, and at the end of the study they commented that the task was actually not so difficult and that there was quite a lot of information available to them in order to perform these inferences.

Participant	**Time**	**Lessons**	**Inferences**
1	26:25	5	9
2	30:44	2	9
3	32:50	3	6
4	42:17	6	15
5	40:15	5	7
6	31:30	6	7
7	48:00	3	6
8	37:12	4	8
9	31:03	3	7
10	42:45	6	11
Average	**36:18**	**4.3**	**8.5**
Total			**85**

Table 5.1: Basic statistics

In table 5.1 we can see some basic statistics for the results of the study. The time devoted by each participant was 40 minutes, but some of them finished interacting with the system somewhat sooner, while others took much longer. On average, the time participants spent with the system was around 36 minutes. In this time, an average of 4 lessons were covered, and 8.5 inferences were made per participant.

Although the number of inferences per participant is not very high, we can see that in total we collected a total of 85 inferences. In the following sections we present these inferences. In section 5.4.1 we present in detail two excerpts from comments made by one of the participants, and the inferences derived from them. In section 5.4.2 we present in a more concise way the motivation diagnosis knowledge rules inferred from this study.

5.4.1 Two excerpts in detail

In this section we present two excerpts from the comments made by participant 1 of this study, while viewing the interaction of a student following the instructional path number 5 in the prototype ITS (see chapter 4).

Excerpt 1

```
Participant:
```

```
OK, well that was interesting 'cause he seemed to show
hesitation at the start and the end, [...] But the middle ones he
completed very quickly in general. I would say that this reinforces my
belief that he is interested and confident, and he put effort into
it. I would be inclined to say that he is satisfied, I still don't
think that you can say very much about the sensory interest.
```

```
Interviewer:
```

```
And [...] why do you think he is satisfied at this point?
```

```
Participant:
```

```
Well, because from the movements of the mouse, he is
hovering the mouse over the answers each time, he wasn't randomly
moving the mouse, he is looking for the answer, and obviously thinking
about what the right answer was, [...] and that he didn't take a
long time to answer the questions. To me that would suggest that the
task is interesting enough to complete with some attention and to do
it properly, if you like. [...] So, I would increase the
satisfaction here, just for the fact that he did it with confidence,
```

Inferred rule from Excerpt 1

This excerpt illustrates a motivation diagnosis rule which infers about the satisfaction of the student based mainly on the interaction with the interface. Because the mouse movement through the interface is not at random, the participant could infer that the student was paying attention to the task. Because it was quickly performed, he inferred that he was confident. And given that:

1. He was interested in the task
2. He was confident
3. He performed the task well

the participant could infer that he would be highly satisfied. In figure 5.4 we present this graphically. We also have to note the dashed arrow and box on figure 5.4. This rule comes from another excerpt, but serves to illustrate the complex nature of the knowledge inferred through this study. We see that performing a task quickly can also mean lack of interest, but it is the combination of other evidence that can lead us to believe that in this case a quick performance was due to confidence.

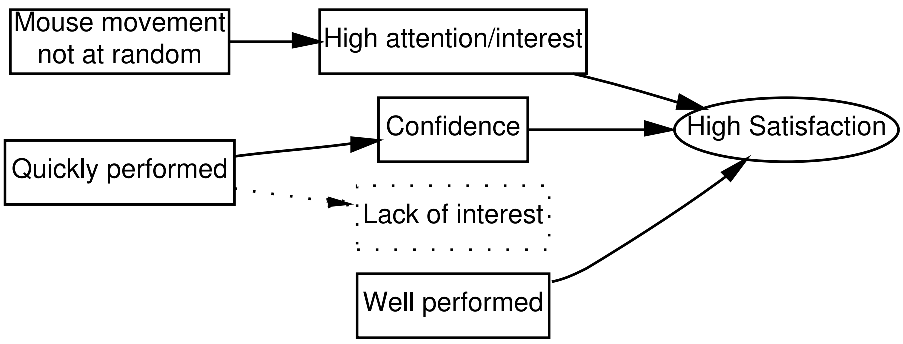


Figure 5.4: Inference rule from Excerpt 1.

Excerpt 2

Participant:

OK, he is filling in the easy ones first, which is a good tactic, I'd also say that's a good indicator of confidence, [...] That he is confident, that he is looking for the ones that he knows first, [...] but he didn't really fill very much in, I was surprised by that.

[...] because he's missed so much out, that would then indicate to me that he doesn't know it. He's given up very quickly, and he's only filled in the ones he knows, just because he cannot pull the rest of the numbers out. I would be worried that his confidence would be decreasing again there, and his effort has obviously gone right down, [...] I would say he's obviously looked at the task, and thought, I don't know this, so just going to do this a little bit, and that's it, but he has less than half right, so...

[...] I would be worried that his satisfaction might decrease there [...] Just based on the fact that he hasn't got these right, [...] I would have a look at his feedback, first, before I reduce it, because I think his satisfaction might depend on what feedback is given, but I would be inclined to say that it would deteriorate there.

Interviewer:

The system says [...] 'Not too bad. Would you like to continue?'

Participant:

He hasn't been marked right for 'ju-san' he has only missed out the hyphen, [...] I think that his satisfaction might be decreased by that, certainly from that [...] By the fact that he was actually really close to the solution, but he has been marked 'wrong'? [...]

```
I'm sure that the hyphen is important in Japanese, but in terms of a
beginner, that is a very minor error, and he got the important part,
which is the sound and the rhythm, you know the structure right, so
that's not fair, if you like...
```

Inferred rule from Excerpt 2

We can see that this excerpt is more complex than Excerpt 1, as there are more factors that influence the final inference. As before, we represent graphically the inference from this excerpt in figure 5.5.

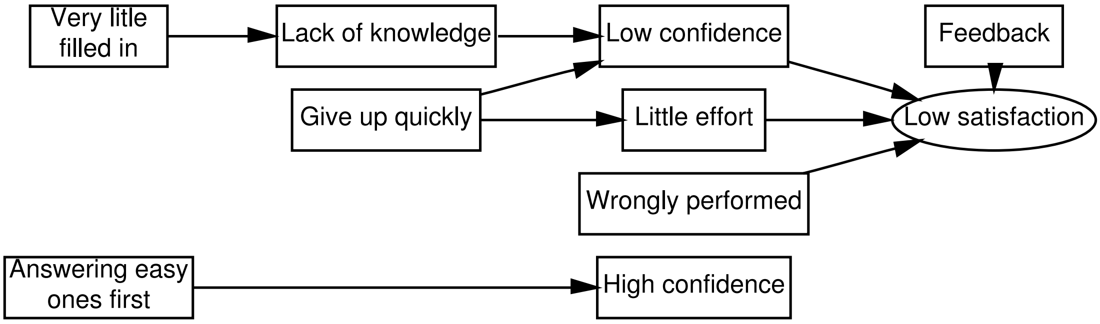


Figure 5.5: Inference rule from Excerpt 2.

The first thing to note is that this time we have conflicting inferences from the same excerpt. Thus, the participant inferred firstly that the student would be highly confident as he was being selective on which exercises he was performing, and trying to solve first the ones that he knew. But later the participant changed his mind, as the student completed only a small part of the lesson, and he gave up very quickly. This led the participant to infer that the student's confidence would be low. Given the new evidence, the participant inferred that the student would be dissatisfied based on:

1. His low confidence;
2. the little effort that he put in the task; and
3. the fact that the task was wrongly performed.

But it is also interesting to see how the feedback given to the student would also have an important influence in the student's satisfaction. The participant suggests that the student's satisfaction is likely to decrease, but she prefers to wait for the actual feedback given by the system to comment about her final inference on the student's motivation. The fact that the feedback is somehow negative reinforces the participant's opinion that the student's satisfaction will decrease. A more appropriate feedback would have avoided a sharp decrease in the student's motivation. Thus, we see that feedback is also an important influence in a student's motivation.

5.4.2 Elicited motivation diagnosis knowledge

First approximation

By analysing the rest of the data gathered from this study we first elicited a large number of provisional motivation diagnosis rules, which we present in detail in appendix B. In this section we just present some of them, due to lack of space.

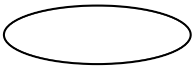
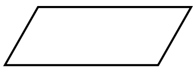
Node	Description
	Steps where the student's motivational models are involved
	Steps where interface issues are involved (e.g. moving the mouse a lot)
	Performance issues (e.g. time required to perform exercise)
	Other intermediate steps (e.g. student finds exercise harder)
	Steps involving feedback to be given to student
A → B	*A* implies *B*
A, B → C	*A* AND *B* imply *C*

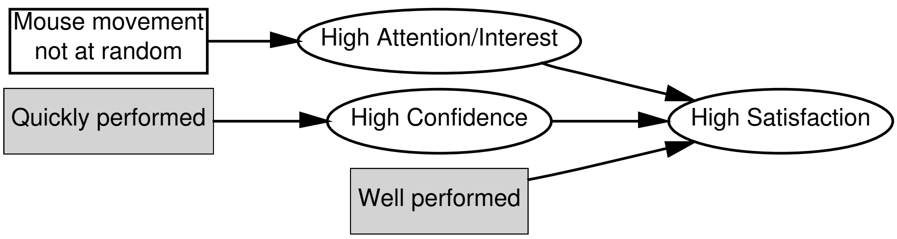
Table 5.2: Graphical conventions for table 5.3.

By carefully analysing the interviews with the participants, we created a provisional set of rules, some of which can be seen in table 5.3[3]. In this table we have attempted to be true to participants' comments and therefore the rules are represented midway between the verbose explanations of participants and the necessary formalisation for their implementation in an Affective Tutor.

In order to make table 5.3 easier to understand, we have followed the graphical conventions presented in table 5.2 in order to represent the different parts that make each rule. These conventions are not always exclusive, and it could be argued that some of the rule parts could be equally classified as some other type of node. This is mainly the case for the nodes representing "interface issues" and "peformance issues". For example, the item "Answered easy ones first" in rule (f) in table 5.3 is categorized as "interface issues", but it is clear that it does also represent performance issues. Nevertheless, the emphasis of this item seems to be on the order in which the exercises were filled in, and as such, we categorized it as "interface issues". The same procedure was followed for all items, classifying them according to the main emphasis of the item, even though another classification could have been possible. In any case, the type of node used to represent each part of a rule serves only as a guide to understand the type of issues involved in each rule, but does not have any importance in order to elicit the final set of rules.

Table 5.3: Examples of motivation diagnosis rules

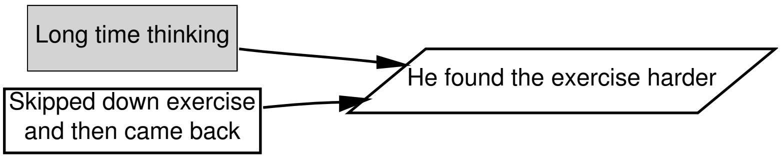
(a)



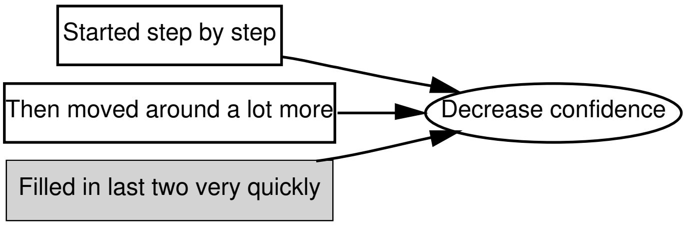
[3]The complete set of provisional motivation diagnosis rules can be found in section B.3.

Table 5.3: Examples of motivation diagnosis rules (continued)

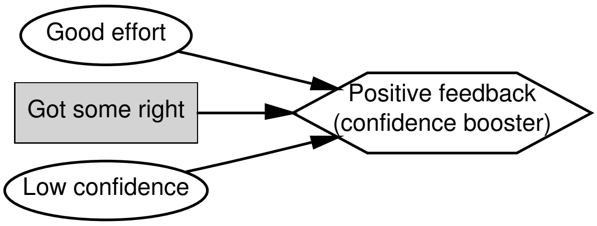
(b)



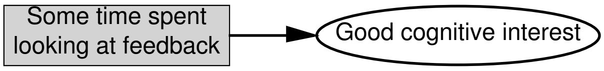
(c)



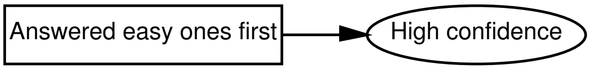
(d)



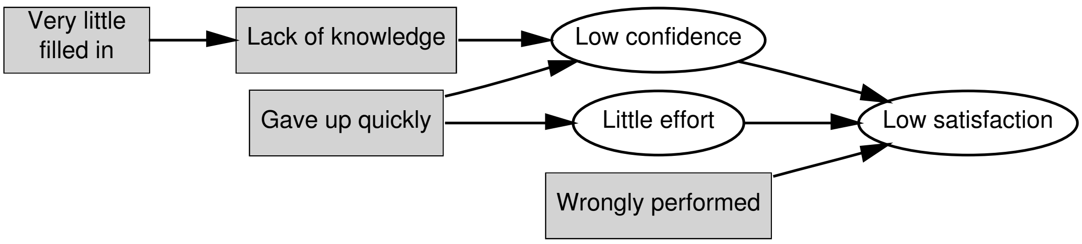
(e)



(f)



(g)



(h)

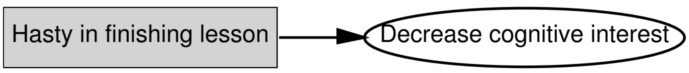
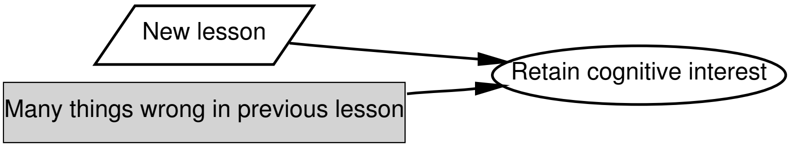


Table 5.3: Examples of motivation diagnosis rules (continued)

(i)



As we can see in table 5.3, the inferences of the student's motivational state are based on a variety of factors (e.g. speed of performance, mouse movements, etc.). In figures 5.6 and 5.7 we can see how often each of these factors were mentioned by the participants.

Figure 5.6 shows the number of inferences made by input categories. That is, the number of inferences where a factor of a particular category was used as input to the inference rule. For example, in table 5.3 we can see that there were two rules, (a) and (c), where *mouse movements* was used as an input factor. In figure 5.6 we can see that this category (*mouse movements*) was mentioned as an input factor in six of the rules elicited from this study.

Similarly, figure 5.7 shows the number of inferences made by output categories. For example, in table 5.3 we can see that there were four rules, (a), (c), (f) and (g), where *confidence* was used as an output factor. In rules (a) and (g) *confidence* is both an input and an output factor, while in rules (c) and (g) it is only an output factor.

As we can see in figure 5.6, the input factors mentioned more often by the participants during this study were those related to students' performance. The main category in figure 5.6 is that of *characteristics of performance*, which was mentioned in 41 of the 85 provisional rules elicited. This category includes a number of characteristics which relate to the way the student performed during the interaction, such as the order in which he did the exercises, whether he gave up or not, etc.

The second most mentioned broad input category was that of *Teaching materials*, in which we include categories such as the *Difficulty of the teaching materials*, issues regarding the previous *History of interaction*, etc. Although not mentioned as often as

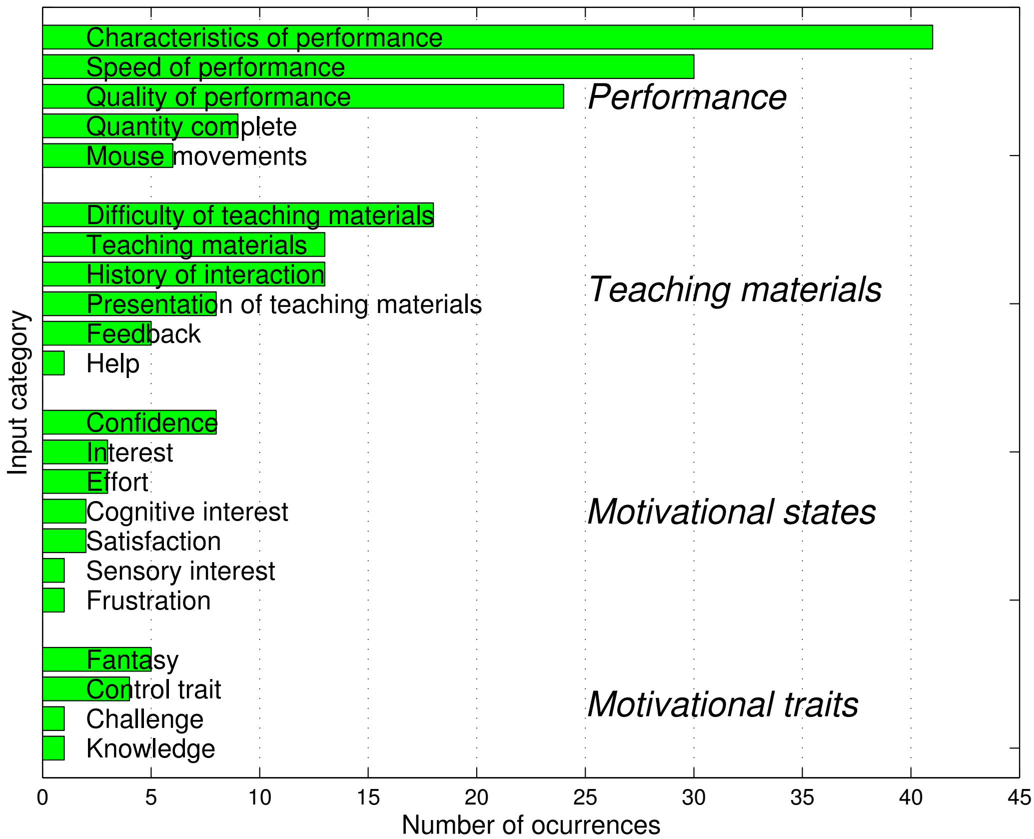


Figure 5.6: Occurrences for each input category.

Performance or *Teaching materials* issues, we can see in figure 5.6 that the student's *Motivation model* and his *Motivational traits* were also considered in a number of occasions as input factors for some of the inference rules.

In figure 5.7 we can see which output categories were mentioned more often by the participants of this study. Not surprisingly, as that was the main purpose of the study, most of the inference rules have as their output a category relating to the student's *Motivational model*, but we can see that there were also a number of cases where the output of some of the rules was in relation to other issues. For example, there were some rules in which the participants inferred about student's knowledge on the subject, about the feedback that should be provided to the student, etc.

In the broad category of *Motivational model*, we see that the factor that seemed to

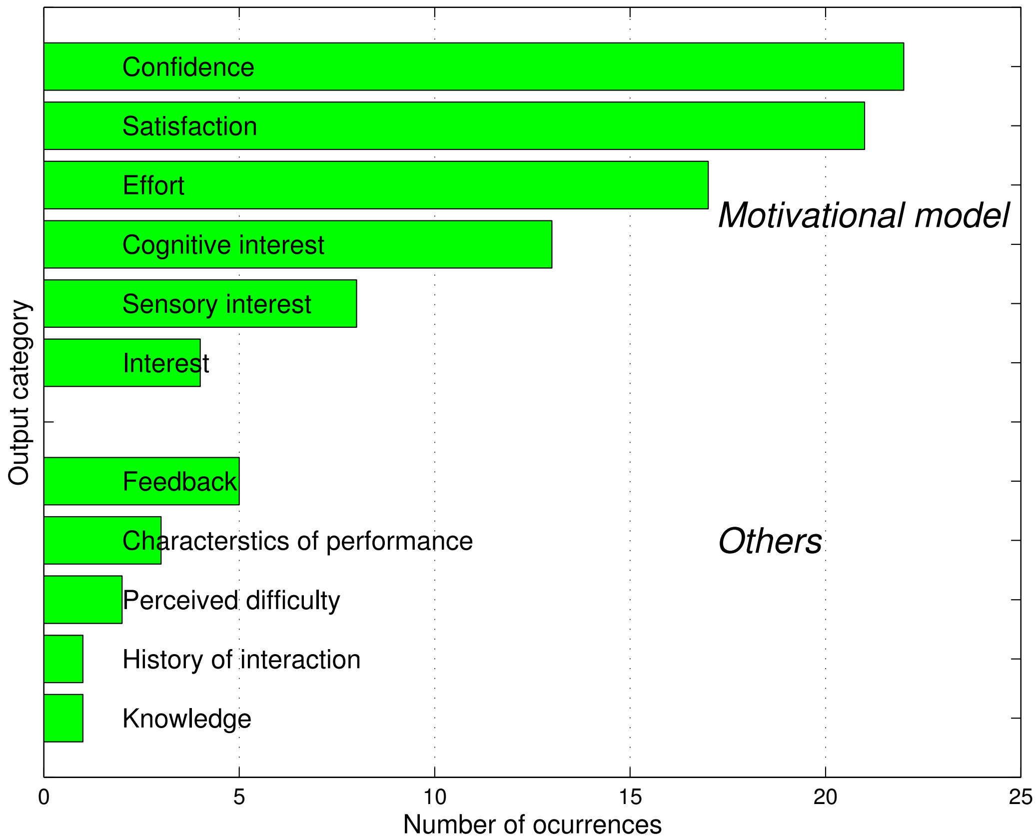


Figure 5.7: Occurrences for each output category.

be the easiest for the participants to infer was that of *Confidence*. The factor *Interest* in figure 5.7 represents comments where the participants did not explicitly specify what type of interest (i.e. cognitive or sensory) they were referring to, although most of the time we can conclude that they implied *cognitive interest*. Therefore we can see that the factor that seemed more difficult to infer was that of *Sensory interest*.

It is interesting to note here the differences with the data obtained in the self-report study. As we saw in table 4.3 in page 58, the self-report factor most frequently used was that of *Confidence*. As mentioned earlier, Briggs *et al.* (1996) claimed that self-confidence is a factor that is easy to report. As we can see in figure 5.7, it seems that it is also a factor easy to infer.

However it is interesting to see that the participants of this study seemed to find it harder to infer about a student's effort than about their satisfaction. On the other hand, the participants of the self-report study seemed to show the opposite behaviour, so that they would report their effort on more occasions that their satisfaction. A similar situation arises with the factors *Cognitive interest* and *Sensory interest*. *Cognitive interest* seemed to be easier for participants of this study to infer, while participants of the self-report study seemed to find it easier to report their *Sensory interest*.

Final set of motivational diagnosis rules

By carefully analysing the provisional rules, we created a final set of rules, which can be seen in tables 5.4 to 5.13. In order to create this final set of rules, we removed repeated rules, very similar rules and rules whose formalisation was not possible because they were not properly reasoned (for example, one of the comments made by one of the participants: "His satisfaction would be high because, well, he obviously looks satisfied"!). At the same time, we decomposed the rules with intermediate outcomes into different rules. For example, the provisional rule (a) in table 5.3 transforms into final rules IS1 ("Satisfaction high when the quality, the confidence and the cognitive interest were high" and IC2 ("Confidence high when the quality is high and the hesitation is low"). Thus, all the final rules in tables 5.4 to 5.13 have only one final outcome and no intermediate ones.

In total there are 61 rules divided into 10 tables. For ease of reference, we have divided the rules according to the motivational factor that they refer to. At the same time, for each of the motivational factors we have divided the rules according to whether the diagnosis is to increase or to decrease the value of that particular factor. Thus, table 5.4 presents all the rules that infer a high value for the *Satisfaction* motivational factor. Table 5.5 presents all the rules that infer a low value for the *Satisfaction* motivational factor. The rest of the tables are divided similarly.

All the tables from 5.4 to 5.13 follow similar conventions, which we describe below by explaining table 5.4. Table 5.4 presents the rules that infer a high value for the *Satisfaction* motivational factor. Each rule has a reference code in the first column of the table, ranging in this case from *IS1* to *IS9*. The letters represent an abbreviation of the type of rule, in this case *Increase Satisfaction*. The last column of the table

Rule	pre(Quality)	Quality	Speed	pre(Difficulty)	Difficulty	Control	Feedback	Satisfaction	Confidence	Effort	Cognitive interest	Control	Challenge	Output	Based on
	PERFORMANCE			*TEACHING MATERIALS*				*MOTIVATION MODEL*				*MOT. TRAITS*			
IS1		High							High		High			High	1a
IS2	Low	High	Fast											High	2a
IS3		High					Enc							Inc	10a
IS4		High					Enc				High			High	10j
IS5											High			High	10i
IS6		High		X	>X					High			High	High	7e
IS7		High	Fast	X	>X									Inc	8d
IS8				X	<X			Low	Low					Inc	3d
IS9						High						High		High	4n

Table 5.4: Increase Satisfaction diagnosis rules

Rule	Quality	Speed	pre(Give up)	Give up	pre(Difficulty)	Difficulty	Control	Fantasy	Feedback	Confidence	Effort	Control	Fantasy	Output	Based on
	PERFORMANCE				*TEACHING MATERIALS*					*MOT. MODEL*		*MOT. TRAITS*			
DS1	Low									Low	Low			Low	1g
DS2	Avg			Yes										Low	6c
DS3			Yes	Yes	X	>X						High		Low	6f
DS4		Very slow												Dec	7a
DS5								Low	Plain				High	Dec	7c
DS6	Very low							Very low						Very low	8g
DS7							Very low					High		Dec	10c

Table 5.5: Decrease Satisfaction diagnosis rules

Rule	PERFORMANCE							TEACHING MATERIALS		Output	Based on
	pre(Quality)	Quality	Quantity	Speed	pre(Give up)	Give up	Hesitation	Difficulty	Feedback		
IC1	Low	High		Fast						High	2a
IC2		High					Low			High	3f
IC3					Yes					Avg	6a
IC4		High		Avg				High		High	7f
IC5			High			No				Inc	9g
IC6		High							Enc	Inc	10a
IC7		High		Avg						High	10f

Table 5.6: Increase Confidence diagnosis rules

Rule	Quality	Quantity	Speed	pre(Give up)	Give up	Mouse movements	Hesitation	Performed in order	pre(Difficulty)	Difficulty	Control	Output	Based on
	PERFORMANCE								*TEACHING MAT.*		*MOT. TRAITS*		
DC1						High	High					Dec	1c
DC2							High					Low	2h
DC3							High	No				Low	4d
DC4					Yes							Low	4k
DC5				Yes	Yes				X	>X	High	Low	6e
DC6		Avg					High					Low	8c
DC7			Very slow				High					Very low	8e
DC8	Very low								X	>X		Very low	8h
DC9		Very low							X	>X		Dec	9a

Table 5.7: Decrease Confidence diagnosis rules

Rule	Quality	Quantity	Speed	pre(Give up)	Give up	Hesitation	Performed in order	Difficulty	Output	Based on
	PERFORMANCE							*TE. MAT.*		
IE1					No	High			Inc	3a
IE2		Very high							High	5a
IE3		High			No				High	5b
IE4							Yes		Inc	4g
IE5				Yes					High	4o
IE6		High	Avg					High	High	6d
IE7	High		Slow						High	10h

Table 5.8: Increase Effort diagnosis rules

Rule	Quality	Quantity	pre(Speed)	Speed	Give up	pre(Difficulty)	Difficulty	Output	Based on
	PERFORMANCE					*TEACHING MATERIALS*			
DE1				Fast			High	Low	2c
DE2		High						Avg	4f
DE3					Yes			Low	4l
DE4	High	Avg			Yes			Low	8a
DE5	Avg	Low			Yes			Low	8f
DE6	Low	Low				X	>X	Dec	9b
DE7			Y	<Y		X	>X	Dec	9d

Table 5.9: Decrease Effort diagnosis rules

Rule	pre(Quality)	Speed	Help	Hesitation	Feedback speed	pre(Difficulty)	Difficulty	Satisfaction	Confidence	Output	Based on
	PERFORMANCE					*TE. MAT.*		*MOT. MODEL*			
ICI1					Slow					High	1e
ICI2						X	<X	Low	Low	Inc	3e
ICI3			Yes							Inc	4j
ICI4		Slow		Low						High	10k
ICI5	Low									Avg	1i

Table 5.10: Increase Cognitive Interest diagnosis rules

Rule	Quality	Quantity	Speed	Hesitation	pre(Difficulty)	Difficulty	Output	Based on
	PERFORMANCE				*TEACHING MATERIALS*			
DCI1			Very fast				Dec	1h
DCI2	Low			High			Dec	5f
DCI3			Fast			High	Low	4c
DCI4	Low	Low			X	>X	Dec	9c

Table 5.11: Decrease Cognitive Interest diagnosis rules

Rule	Reset game	Mouse movements	pre(Speed game)	Speed game	Feedback	Output	Based on
	PERFORMANCE		*TEACHING MATERIALS*				
ISI1		Fast				Inc	2e
ISI2	Yes					High	5c
ISI3			X	>X		High	5e
ISI4					Yes	Inc	4m

Table 5.12: Increase Sensory Interest diagnosis rules

Rule	Sensorial	Fantasy	Output	Based on
	TEACHING MATERIALS	*MOT. TRAITS*		
DSI1	Very low		Low	2d
DSI2	Low	High	Low	4e

Table 5.13: Decrease Sensory Interest diagnosis rules

gives the provisional motivation diagnosis rule from which this rule was inferred. For example, *IS1* was inferred from provisional rule *1a*[4]. The rest of the columns represent the actual rules.

Each column represents an input factor, which are divided into the same broad categories as in figure 5.6. Not all the tables representing the Motivational Diagnosis rules have the same columns, as not all the rules inferred make use of the same input factors. For instance, in table 5.4 the motivational traits that are used as input for some of the rules are *Control* and *Challenge*. In table 5.5 the motivational traits are *Control* and *Fantasy*. On the other hand, some of the rules make no use at all of motivational trait factors (for example, table 5.6), so these are not present on the corresponding tables.

The input factors for both *Performance* and *Teaching Materials* (except those starting with *pre*) refer to characteristics of the current instructional unit (i.e. the one being performed or just finished). For example, rule ISI1 in table 5.12 would mean that if a student is moving the mouse quickly while studying a lesson, we could infer that his sensory interest would increase.

The input factors starting with *pre* refer to characteristics of the instructional unit immediately before the current one. For example, rule IC1 in table 5.6 would mean that if a student had performed the last lesson poorly but did the current one well and fast, we could infer that his confidence would be high.

In greater detail, the factors given as input for the rules and the possible values that they can take are:

- Performance
  - **Quality**. Correctness of the answers provided to the current instructional unit.

    Values: Very low, Low, Avg (i.e. Average), High, Very high. The values range from Very low (i.e. a very low correctness of the answers provided) to very high (i.e. a very high correctness). A similar meaning is applied to all the factors whose values fall in this range.

[4] As mentioned earlier, the complete set of the provisional rules is given in appendix B, section B.3.

- **Quantity**. Percentage of answers attempted.

  Values: Very low, Low, Avg, High, Very high.

- **Speed**. Time spent in doing the instructional unit.

  Values: Very slow, Slow, Avg, Fast, Very fast, X >X, X <X. The values range from Very slow (i.e. the student took a long time to perform the lesson), to Very fast (i.e. the student took little time to perform the lesson). At the same time, when used in conjunction with the factor pre(Speed), we can have the values "X >X" (i.e. the student performed this lesson slower than the previous one) and "X <X" (i.e. the student performed this lesson faster than the previous lesson). A similar meaning is applied to all the factors whose values fall in this range.

- **Give up**. Whether the student chose to give up the lesson or not.

  Values: Yes, No. Whether the student gave up this lesson or not.

- **pre(Quality)**, **pre(Quantity)**, **pre(Speed)** and **pre(Give up)**. Same as the equivalent factors above, but referring to the previous lesson instead of the current one.

  Values: Same as the equivalent factors above.

- **Mouse movements**. Speed of mouse movements.

  Values: Very low, Low, Avg, High, Very high

- **Hesitation**. Degree of hesitations showed by the student.

  Values: Very low, Low, Avg, High, Very high

- **Performed in order**. Whether the exercises were performed in the order provided or not.

  Values: Yes, No

• Teaching Materials

- **Difficulty**. Level of difficulty of the current exercise.

  Values: Very low, Low, Avg, High, Very high, X >X, X <X.

  - **pre(Diff)**. Level of difficulty of the previous exercise.

    Values: Same as Difficulty.

  - **Control**. Level of control available in the current lesson.

    Values: Very low, Low, Avg, High, Very high.

  - **Fantasy**. Degree of fantasy in the current lesson.

    Values: Very low, Low, Avg, High, Very high.

  - **Feedback**. Characteristics of the feedback provided.

    Values: Plain, Enc. "Plain" feedback refers to feedback without any motivational content. "Enc" feedback refers to encouraging feedback, which tries to emphasize the positive aspects of the student's performance in order to encourage him to continue.

- Motivation model

  - **Satisfaction**, **Confidence**, **Effort**, **Cognitive Interest** and **Sensory Interest**. Value of the corresponding factor in the student's motivational model.

    Values: Very low, Low, Avg, High, Very high.

- Motivation traits

  - **Control**, **Fantasy** and **Challenge**. Value of the corresponding motivational trait.

    Values: Very low, Low, Avg, High, Very high.

The output of the rules refers to the detection of the given motivational factor under the circumstances given by each of the rules and can take the values: Very low, Low, Avg, High, Very high, and Dec (i.e. Decrease), Inc (i.e. Increase). Whehter an output value is in the range *Very low* ... *Very high* or in the range *Dec* ... *Inc* relates to the elicitation made by the participants.

For example, rule IS4 can be expressed as: "If the student performs correctly most of the lesson; he was given encouraging feedback; and we know that his cognitive interest is high; then we can infer that his satisfaction will be high".

Rule IS7 can be expressed as: "If the student performs correctly most of the lesson; he performs it fast; and the difficulty of the current lesson is less than the difficulty of the previous one; then we can infer that his satisfaction will increase over its present value".

As we can see from tables 5.4 to 5.13, the number of motivation diagnosis rules inferred from this study was considerable and they provide an empirically motivated approach to the issue of motivation diagnosis. In the following section we discuss some of the issues raised by these results.

5.5 Discussion

While far from being the definitive answer to motivation diagnosis in affective tutoring systems, this study offered us important clues as to which aspects of the instruction seem to be the most relevant in order to detect students' motivational state, and it provided us with a considerable set of motivation diagnosis rules.

As mentioned earlier, participants in this study were initially quite convinced that the task would prove extremely difficult and that it would be virtually impossible for them to extract any useful information about the student's motivational state without being able to see him. But despite the original doubts of most participants, we have seen that we were able to infer a large number of motivation diagnosis inference rules.

More importantly, by only showing them the student's interaction with the tutoring system, these rules are based on very concrete aspects of the interaction, such as mouse movements, quality of performance, etc., which can be easily detected in a tutoring system. On the other hand, we believe that if the participants had been able to see the student himself, many of the inferences about his motivational state would have been based on their gestures, posture, etc., which would prove much harder to detect in a regular tutoring system.

In conclusion, we can say that the results of this study showed that it is feasible to infer motivation diagnosis knowledge based only on the information provided by the computer interaction with a tutoring system. We managed to gather a considerable number of motivation diagnosis rules, but these have to be validated, which we discuss in the following chapter.

Chapter 6

Motivation Diagnosis Validity Study

6.1 Introduction and goal of the study

As a result of the Motivation Diagnosis Study described in chapter 5, we obtained a number of motivation diagnosis rules, which were described in section 5.4.2, but whose validity remained to be analysed.

Cross-participant comparison was not an appropriate way to validate the given set of rules, as the number of rules elicited was not large enough to provide many groups of rules which could be applied under the same conditions. Similarly, comparison with the self-report study presented in chapter 4 was not appropriate because there was no reason to believe that the self-report is necessarily accurate, as 'false' readings can be given under certain circumstances. For example, if the student is too engaged, he would probably forget to update the motivational model. Also, it is likely that students will attempt to 'please' the tutoring system by providing artificially positive readings of their motivation (Reeves and Nass, 1998).

Therefore, we evaluated these rules by performing another study in which participants were presented with an instructional interaction context and were asked to rate the rules that could be applied under those conditions. This study gave us a chance to find which rules from the current set are generally accepted as valid, and which ones are not. We describe this study in this chapter.

6.2 Materials

The validation study took the form of a simple questionnaire which was administered on-line. A question was prepared for each of the 61 rules presented in tables 5.4 to 5.13 in section 5.4.2. Each question took the form of a description of an instructional setting and a question regarding a motivational factor. As an example, we can see figure 6.1, which represents the question corresponding to the rule IE6 in table 5.8. A complete list of all the questions can be seen in section C.2.

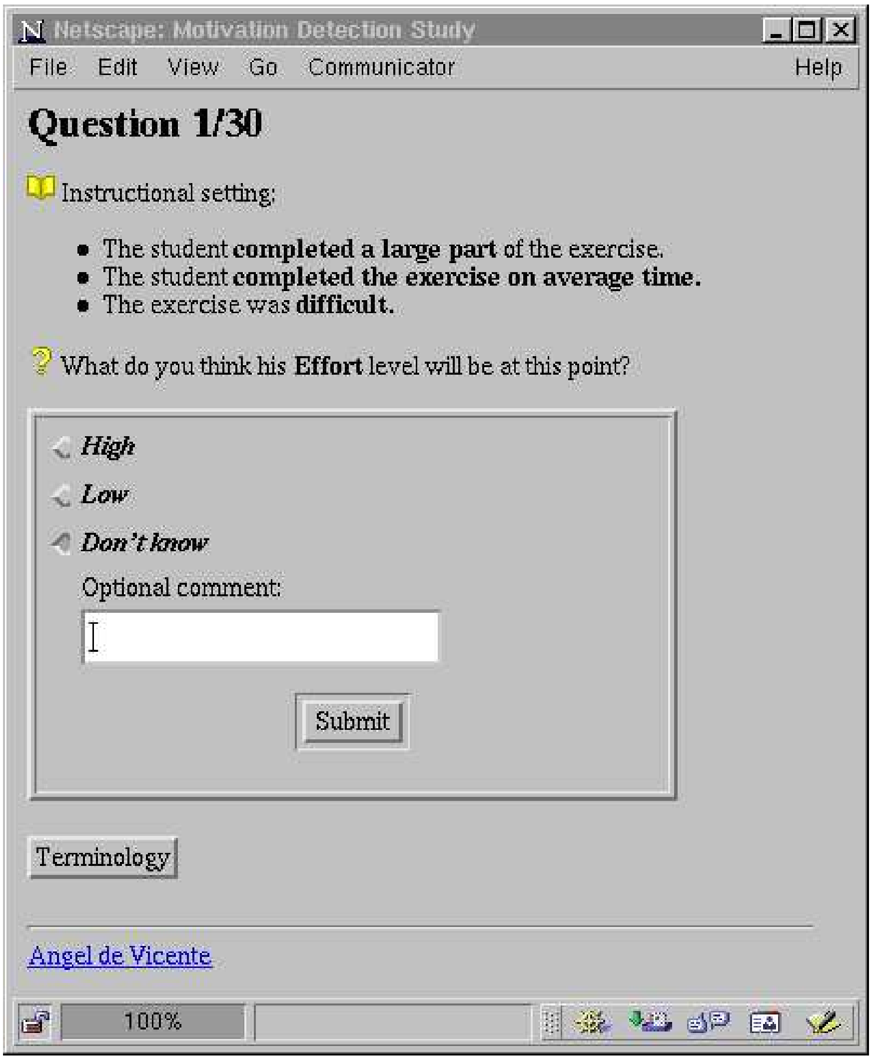


Figure 6.1: Validity Study Sample question.

Each participant was asked 30 questions. For each question, the instructional setting was given as a number of simple sentences representing the situation described by the corresponding rule. The motivation question related to the motivational factor of the given rule. For each question we asked the participants to infer what the value of the motivational factor would be. As possible options we gave them three choices:

- The value given by the corresponding rule.
- The opposite value.
- Don't know.

The *Don't know* option was always the third item in the list. The other two options were ordered in decreasing value. This way the value given by the corresponding rule would be sometimes the first item in the list, and sometimes the second.

For example, rule IE6 in table 6.1 was given to participants with the instructional setting described as:

- The student **completed a large part** of the exercise.
- The student **completed the exercise on average time**.
- The exercise was **difficult**.

Rule	**Quantity**	**Speed**	**Difficulty**	**Output**	**Based on**
	PERFORMANCE		*TEACHING MATERIALS*		
IE6	High	Avg	High	High	6d

Table 6.1: Rule IE6 (Increase Effort) (*Reproduced from table 5.8).*

As we can see, each of the sentences representing the instructional setting correspond to one of the rule's inputs. The motivation question referred to the motivational factor **Effort**, and it was presented to the participants in the following way:

- What do you think his **Effort** level will be at this point?
  - High
  - Low
  - Don't know

In the case of the rule IE6, the output value of the rule is *High*. Thus, the choices for responding to the IE6 question that we gave to the participants were *High*, *Low* (the opposite), and *Don't know*. The participants were asked to choose the option *Don't know* if they thought they didn't have enough information to choose between the other two values. At the same time, they were encouraged to provide any optional comments if they wished to.

6.3 Participants

To obtain participants for this study, we asked for collaboration in a number of mailing lists dealing with technology and education. The prerequisite for participating in the study was to have at least two years of teaching experience and, as a result of this, 33 participants volunteered to take part in this study. In figure 6.2 we see the distribution of years of teaching experience amongst the participants. As we can see, some very experienced teachers took part in this study. The subject area was not a precondition, and therefore there was an ample range of subjects taught by the participants, amongst others: Programming, Maths, Spanish, Biology, Linguistics, Social Studies, Education and Chemistry.

6.4 Methodology

As mentioned earlier, the study was administered as an on-line questionnaire. After a participant agreed to participate in the study, she was given the web address of the instructions for the study. These instructions can be seen in appendix C, section C.1.

In order to avoid duplication of answers, the questionnaire web pages were protected by a username and a password, which was given to each participant after they

Figure 6.2: Teaching experience of participants in Validity Study.

had agreed to participate in the study. The username and the password were valid only once, so that participants could not perform the study more than once. After the participant had read the instructions and provided her username and password, she was directed to the questionnaire proper.

Each participant was asked 30 questions, chosen randomly amongst the 61 available. They were asked to choose the option that they thought would be applicable under the given instructional setting. If they did not have enough information to make a choice, they were asked to choose the *Don't know* option. Regardless of the option they selected, they were also offered the possibility of giving any extra comments and/or qualifications to their answer. Once they had finished answering all the questions, they were given the option to see the cumulative results.

6.5 Results

In total we collected 973 votes (instead of the maximum 990 possible votes), as some of the participants experienced technical difficulties that forced them to finish the questionnaire before they had completed the 30 questions.

As the questions were chosen randomly for each participant, not all the questions got the same number of replies, but all of them got at least 15 replies. For an example of the replies obtained see table 6.2 (the complete list of the results can be seen in table C.1 in page 288). The columns represent respectively: the name of the rule (see tables 5.4 to 5.13); the output value of the rules; the number of participants who answered the question corresponding to that rule; the value of the first choice given to participants; the number of participants that selected that choice; and the same for the second and third options.

Rule	Output	n	Option 1		Option 2		Option 3	
DS1	Low	15	High	0	Low	15	Don't know	0
DS4	Decrease	16	Increase	3	Decrease	2	Don't know	11
DC2	Low	15	High	1	Low	10	Don't know	4
DE1	Low	16	High	6	Low	6	Don't know	4
DE2	Average	17	High	8	Average	0	Don't know	9
ICI4	High	17	High	5	Low	6	Don't know	6
DCI4	Decrease	16	Increase	2	Decrease	10	Don't know	4

Table 6.2: Sample motivation diagnosis rules results.

6.5.1 Was it easy to select an option?

Each rule offered the participants three options: the output value of the corresponding rule, its opposite and *Don't know*. A choice of *Don't know* generally meant a lack of information on which to make a choice (more details in section 6.5.3). In figure 6.3 we can see the distribution of *Don't know* replies amongst the 31 participants that completed the 30 questions.

We can see that some of the participants chose *Don't know* for as many as 20 questions, but that was not usual. On average, each participant chose the option *Don't know* for only 6.94 questions, so most of the time participants seemed to have enough contextual information to make a choice between the other two options.

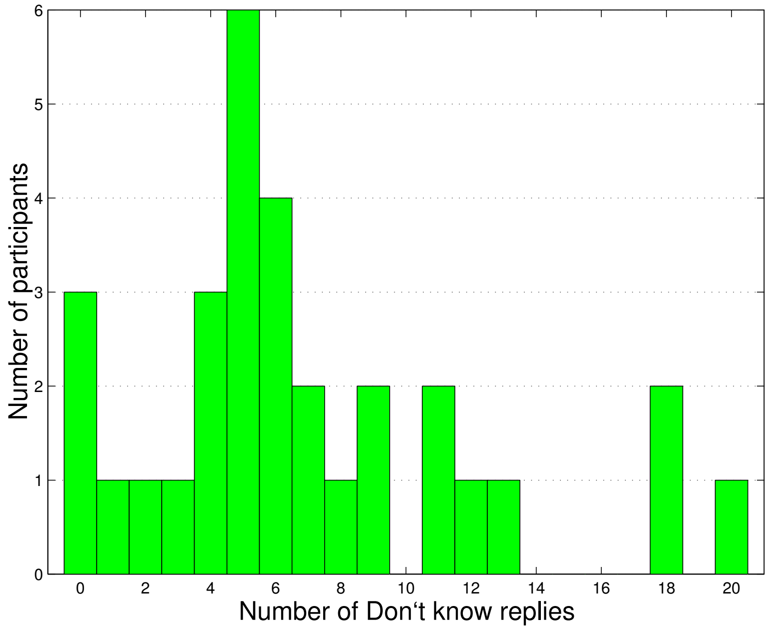


Figure 6.3: Distribution of *Don't know* replies (for the 31 participants that completed the 30 questions).

6.5.2 Which rules to accept?

As mentioned earlier, the main goal of this study was to validate the rules from the Motivation Diagnosis study, so that we could decide which rules should be incorporated into our prototype MOODS system.

In order to decide which rules to accept and which ones to reject, we proceeded in the following manner:

1. For each question we collected the values for the three options (see sample in table 6.2).

2. One of the options is the same as the one suggested by the corresponding rule (column *Value* in table 6.2), and we consider this the **Accept** condition. The other two options are considered **Reject** conditions (either the participant inferred the opposite of what the corresponding rule infers, or selected *Don't know*). We added the votes for the two possible *Reject* options (see sample in table 6.3 for the same rules as in table 6.2).

3. For each of the rules we performed a chi-square *goodness of fit* test, where the null hypothesis is that there is no preference for either *Accept* or *Reject* . Thus, under the null hypothesis we would expect 33% of the votes to go into the *Accept* option and 66% of the votes to the *Reject* option. The probability of the distribution obtained in our study given the null hypothesis is given in column *p* in table 6.3. The complete list of the results of the chi-square tests can be seen in table C.2 in page 290.

4. We consider valid rules those in which the participants showed a statistically significant preference (p<0.01) for the *accept* category.

Rule	n	Accept	Reject	p
DS1	15	15	0	<0.00001
DS4	16	2	14	0.07555
DC2	15	10	5	0.00617
DE1	16	6	10	0.72064
DE2	17	0	17	0.00338
ICI4	17	5	12	0.72844
DCI4	16	10	6	0.01286

Table 6.3: Sample chi-square results.

In figure 6.4 we can see the distribution of the rules according to the number of participants that chose the *accept* option. The line represents the expected number of *accept* and *reject* votes if the null hypothesis was true. Each marker represents the actual number of *accept* and *reject* votes for a particular rule[1]. At the same time, we have drawn a circle for the rules that are accepted.

Thus, we can see in figure 6.4 that the rule DS1 obtained 15 *accept* votes, but 0 *reject* votes. This is clearly a significant preference for *accept*, and as such it is drawn with a circle. Rule DC2 obtained 10 *accept* votes and 5 *reject* votes. This is still statistically significant.

[1]In some cases more than one rule has the same number of *accept* and *reject* votes, so the same dot represents all of them.

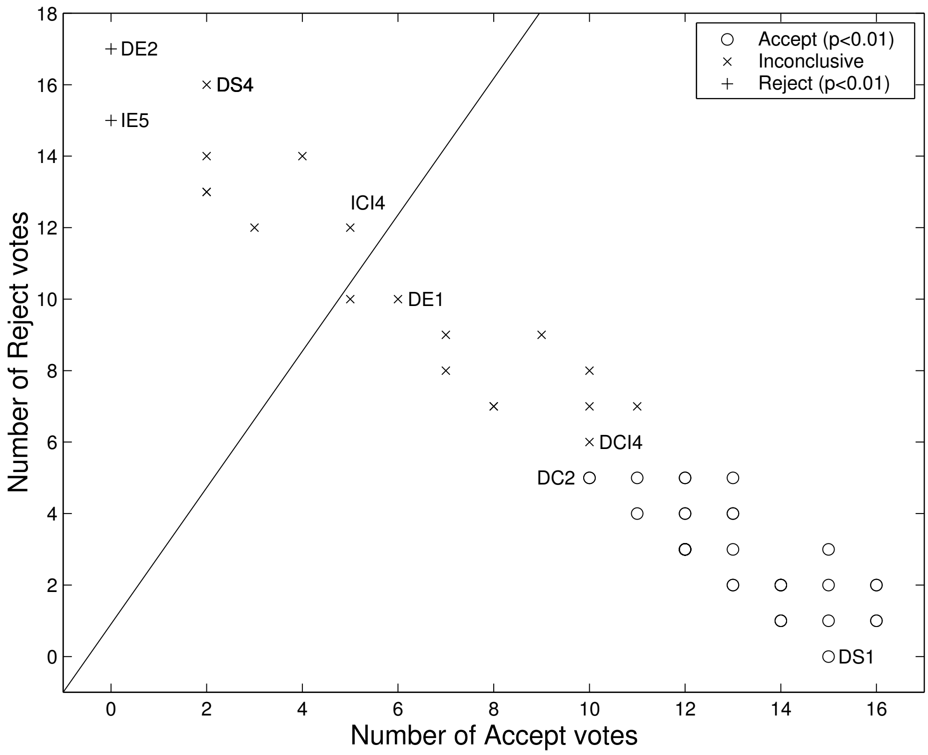


Figure 6.4: Acceptance distribution of Motivation Diagnosis rules.

Rule DCI4 obtained 10 *accept* votes and 6 *reject* votes. This is not statistically significant and therefore we will not accept this rule in our final set. In the case of the rules IE5 and DE2, the distribution obtained has a $p<0.01$, but due to the fact that all the votes were *Reject* votes these rules are not accepted. A complete list of all the rules that we accept, together with the number of *accept* and *reject* votes obtained by each of them is given in table 6.4.

6.5.3 Comments by participants

As mentioned earlier, the participants were encouraged to give extra comments for each of the questions if they wanted to qualify their decision. In this section we briefly analyse these comments.

From the 973 questions answered in total, the participants gave comments to 205 of them. Of these, 130 were given after a *Don't know* vote. We can see in figure 6.5

Rule	n	Accept	Reject
IS1	18	16	2
IS2	15	12	3
IS3	18	16	2
IS4	15	14	1
IS6	17	16	1
IS7	16	14	2
IS9	18	13	5
DS1	15	15	0
DS2	16	12	4
DS3	17	16	1
DS6	15	12	3
IC1	17	13	4
IC2	15	12	3
IC4	15	13	2
IC5	17	12	5
IC6	15	14	1
IC7	17	13	4
DC2	15	10	5
DC3	16	13	3
DC4	15	12	3
DC5	15	12	3
DC6	15	12	3
DC8	16	15	1
DC9	16	14	2
IE1	15	12	3
IE2	16	11	5
IE3	16	11	5
IE6	16	12	4
IE7	16	14	2
DE3	15	14	1
DE4	15	12	3
DE5	18	15	3
DE6	15	12	3
ICI1	17	15	2
ICI3	15	12	3
DCI2	15	11	4
ISI2	15	10	5
ISI3	15	13	2
ISI4	15	12	3
DSI1	16	14	2
DSI2	17	12	5

Table 6.4: Rules to accept. ($p<0.01$)

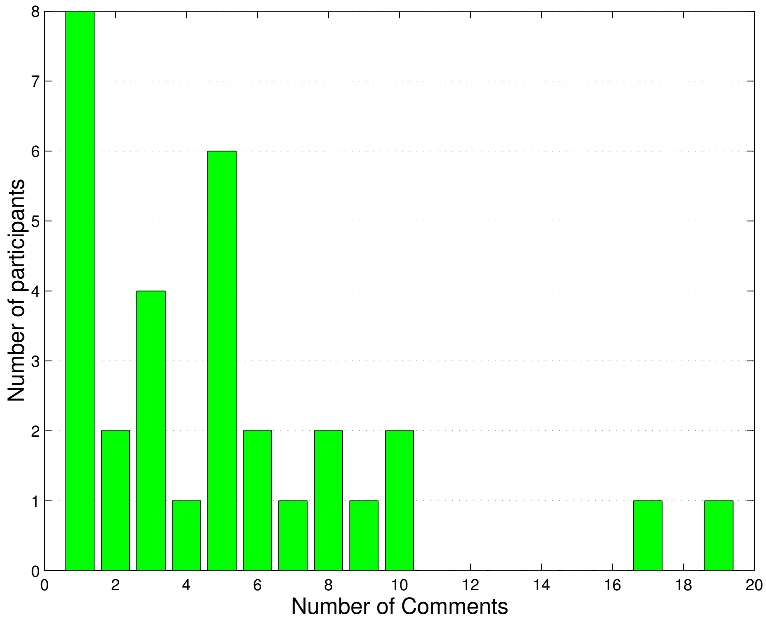


Figure 6.5: Number of comments per participant (for the 31 participants that completed the 30 questions).

the distribution of the number of comments per participant for the 31 participants that completed the 30 questions. We can see that most of the participants made comments to 1 to 5 questions, although two participants gave comments to more than 50% of all the questions.

In many cases the participants commented that there was not enough information in the instructional setting to make an inference for the given motivational factor. But not all comments were of this type: 75 were given after selecting an option other than *Don't know*. These were qualifying comments, which in some cases could help us refine the motivation diagnosis rules.

The comments given by the participants can be divided into a number of categories. We give some sample comments in table 6.5 for each of the main categories and briefly discuss each of them in the following sections.

a) Motivational factor depends on an extra variable. There were many cases when participants were unable to make an inference about the given motivational factor, but

Type	Rule	Output	Participant	Comment
	IS9	High	Don't know	There is no information on whether he actually performed the required tasks.
a)	DS5	Decrease	Don't know	Unclear, since it depends on whether there are enough fantasy characteristics for the student.
	DS4	Decrease	Don't know	Slow work does not indicate degree of satisfaction.
b)	DC7	Very low	Don't know	Some students naturally take long even though they are confident about their abilities.
c)	IS5	High	Don't know	Insufficient information.
	DC1	Decrease	Don't know	Cannot really say with any certainty.
d)	ICI3	Increase	Increase	If help was received and was substantive.
	IS1	High	High	Only if confidence and congitive interest were high.
e)	ICI3	Increase	Decrease	This depends on exactly what help the student needed.
	DCI3	Low	Low	May be low if student is guessing.
	DE2	Average	Don't know	I am going to say low, just because average effort for me would be completion.
f)	IC1	High	Don't know	Confidence would improve after this exercise, but not quite 'High' because of the previous exercise.
g)	ISI1	Increase	Increase	In as much as moving the mouse equates exploration of the interface!
	DE6	Decrease	Increase	Fear of failure, desire to succeed.

Table 6.5: Sample comments.

were able to comment on which other knowledge might be required in order to make such an inference. Examples of this are the first two comments (i.e. comments for rules IS9 and DS5) in table 6.5. The **Output** column refers to the output value of the corresponding rule (details of all the rules were given in tables 5.4 to 5.13). The **Participant** column refers to the option chosen for this rule by the participant whose comment is given in column **Comment**. Thus, we see that one of the participants chose *Don't know* for rule IS9 and commented that she couldn't make a decision because information on whether the student performed the required task was not available.

b) Relation between rule inputs and outputs not clear. The two comments for rules DS4 and DC7 in table 6.5 are examples of another common type of comment made by the participants of this study. For instance, a participant commented for rule DS4 that the speed of performance does not indicate the degree of satisfaction of the student.

c) Not enough information. There were also a number of comments, though less than for categories **a)** and **b)** that were not very informative. These fall in category **c)**, which relates to comments about insufficient information, but without giving any extra information about what other variables would help the participant to make an inference. Examples of this type of comment are given in table 6.5 for rules IS5 and DC1.

d) A choice is made, but under certain assumptions. 15 comments had to do with choices made by the participants, but under certain assumptions. For example, in table 6.5 we can see that a participant inferred the value *Increase* for rule ICI3, but assuming that "help was received and was substantive".

e) A choice is made, but certain information is missing. This category is similar to category **d)**, but in this case an assumption was not made. Rather, the participant pointed out that she was not totally convinced about her choice, which would also be affected by other information. For example, a participant commented for rule ICI3 that the inference should also be based also on "what help the student needed".

f) *Don't know* selected, but a choice is actually made. In a considerable number of occasions, the participants chose the option *Don't know* but at the same time they chose one of the other options in the comments. We can see this for rule DE2, where the participant chose *Don't know*, but she pointed out that she was "going to say low [...]".

g) Elaborations. Quite a large number of comments were simple elaborations on the choice made. See for example the comment given by one participant for rule DE6 in table 6.5, where she pointed out that she chose *Increase* and gave the comment "Fear of failure, desire to succeed" as a way of explaining the reasoning behind her choice.

6.6 Discussion

As mentioned in section 6.1, the goal of this study was to validate the rules obtained in the study described in chapter 5. We obtained sufficient data to perform a *goodness of fit* test for each individual rule, in order to ascertain whether participants preferred the *accept* option over the *reject* ones. As a result we managed to reduce the original set of rules to 41, which were given in table 6.4.

At the same time we gathered a large number of comments from the participants, which could help us to further improve some of the rules. As seen in section 6.5.3, some of the comments made by participants were not very informative (types **b** and **c**). These types of comments helped us to find that the participant thought that the rule was not complete, that some information was missing, but she didn't specify exactly what . Similarly, comments of type **g** were simply elaborations of the choice made, so they are useful to reinforce our confidence in the validity of the rules, but would not be useful to improve the rules.

On the other hand, the comments of type **a**, **d**, **e** and **f** could prove very useful in order to further improve the set of motivation diagnosis rules. It would be interesting to analyse whether the comments made for the non-accepted rules could help us to add/modify/remove some conditions of the rules, so that in a subsequent study we could incorporate these rules in the accepted set. For example, we see in table 6.5

that one of the comments for the rule DS5 was that the inference would depend on "whether there are enough fantasy characteristics for the student". If similar comments were made by other participants about this rule, we could add a condition to test this. In a subsequent study we could test whether adding this condition would make this rule more widely acceptable.

This would be interesting further work, but it is not important for the goal of our validity study. By performing this study we have managed to filter the rules that were not widely accepted by participants. We can only speculate on why some of the rules were not commonly accepted, but there could be a number of reasons, amongst others:

- The lack of educational context meant that participants were not able to understand certain rules, and therefore it was difficult for them to select an option different than "Don't know". Although we have seen that the option "Don't know" was not selected too often, we have also seen that many participants mentioned the need for extra information in certain cases in order to make an inference. It would be interesting to see whether better agreement rates would be achieved if a richer educational context was provided with each rule.

- A participant of the Motivation Diagnosis Study described in chapter 5 had a theory about motivation diagnosis that was not shared by most of the participants in the Validity Study.

- The transcription of some of the rules from the Motivation Diagnosis Study was not accurate. Perhaps a participant of that study did not describe a rule appropriately. Or perhaps we made some mistakes when formalising the rules, for example by omitting some necessary background information.

In this chapter we have presented the empirical study that we performed in order to validate the motivation diagnosis rules presented in chapter 5. By performing this study we filtered the original rules to a smaller set of rules that were commonly accepted by the participants of the study. In chapter 7 we describe how this set of rules was incorporated into the MOODS prototype.

Chapter 7

MOODS. A prototype Affective Tutor Simulation

7.1 Introduction

In chapter 3 we presented an outline of the design of an 'Affective Tutor', but it is in this chapter that we give a detailed account of the design of MOODS, our prototype Affective Tutor Simulation.

MOODS is not a tutoring system, but rather it is a shell where we incorporated all the knowledge regarding motivational planning and motivation diagnosis discussed in previous chapters, but without any actual instructional content. As it is, MOODS is a tool where we can easily evaluate the motivational strategies discussed throughout this dissertation.

We developed an "Affective Tutor" simulation rather than an actual tutoring system due to a number of reasons:

- **Simplicity of development**: as we can see in section 7.4.2, the motivational planning strategies of MOODS require a large number of lessons to be available for selection. In order to develop MOODS with instructional content it would have required us to create a very large number of instructional materials, a time-consuming task which was outside the scope of our work.
- **Simplicity of evaluation**: the evaluation of the motivation diagnosis strategies

implemented in MOODS would have been more difficult with an actual tutoring system than with the developed simulation. As we discuss in chapter 9, an evaluation of the motivation diagnosis strategies implemented in MOODS based on a simulation allowed us to gather more data than it would have been possible with an actual tutoring system.

- **Generality**: MOODS can be seen as the engine of an actual "Affective Tutor", and as such, it is a generic tool that could be used in the development of actual Affective Tutors.

In this chapter we give a detailed account of the design of MOODS, starting with an overview of its overall structure in section 7.2, followed in the remaining sections by details of each of its components. We leave for chapter 8 a demonstration of an interaction with MOODS, and for chapter 9 the evaluation of the strategies implemented in MOODS.

7.2 MOODS Structure Overview

Before we describe each of the individual components of MOODS, we describe in this section its general structure, which can be seen graphically in figure 7.1.

MOODS is developed around a basic instructional cycle composed of three steps, which can be seen inside the thick outlined box at the bottom of figure 7.1:

1. The selection of a lesson.

2. The interaction of the student (Performance).

3. The feedback provided to the student and his response to it.

The rest of figure 7.1 is the actual structure of MOODS, which is designed around the mentioned instructional cycle. There are three main types of components in this structure:

1. Student models, where information about the student is kept. These are represented as oval-shaped dotted boxes in figure 7.1 and are:

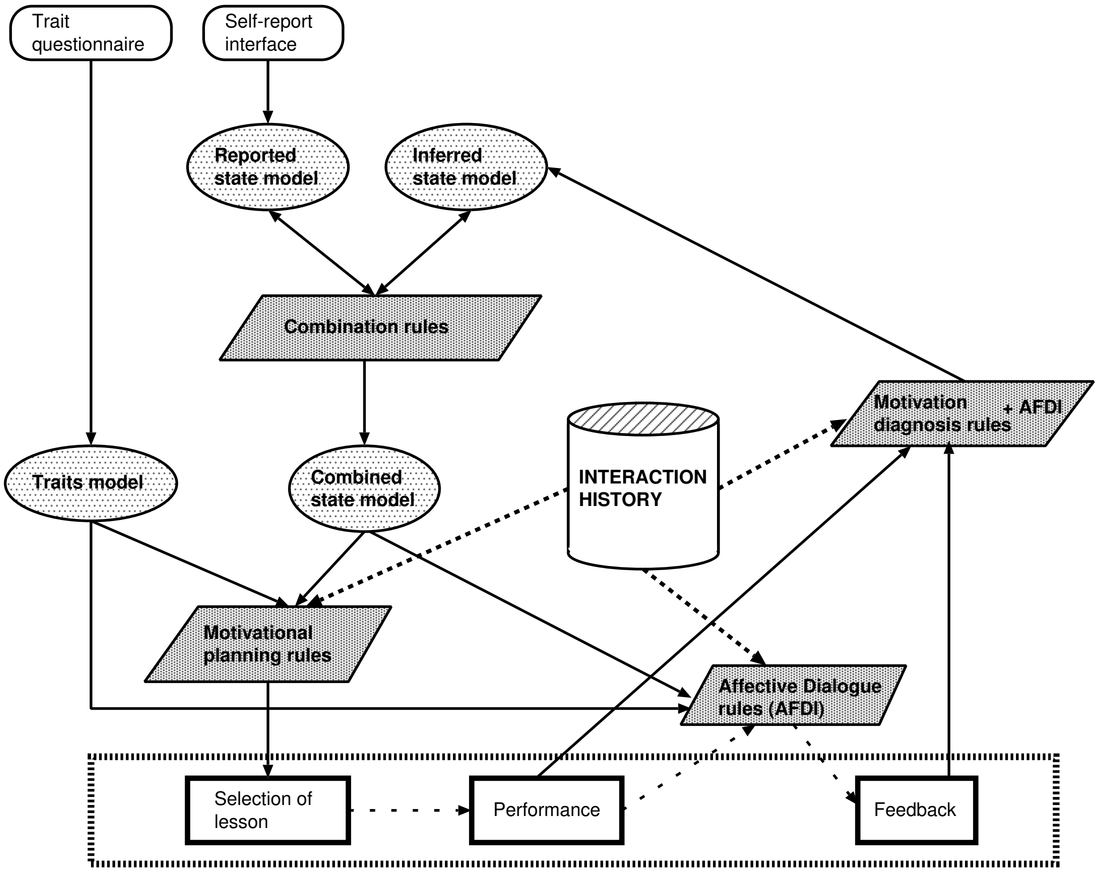


Figure 7.1: MOODS Structure Overview

- **Traits model**: this is a representation of the student's trait characteristics, which can be obtained by a trait questionnaire given to the student at the beginning of the interaction, as discussed in chapter 4.
- **Reported state model**: self-reported motivational model, obtained via the self-report interface, which was discussed in chapter 4.
- **Inferred state model**: inferred motivational model, obtained via the validated motivation diagnosis rules seen in chapter 6.
- **Combined state model**: since the reported state model and the inferred state model do not always coincide, a combined one has to be created, which is considered to be the most accurate model of the student.

2. Knowledge-based rules, that contain the knowledge necessary in order to take decisions about the next lesson to select, what feedback to provide, etc. These are represented as trapezoidal grey boxes in figure 7.1 and are:

    - **Motivational planning rules**: these are the rules that decide on the type and difficulty of the next lesson to be selected. As we can see in figure 7.1, they are informed by the traits and the combined state models and the Interaction History. These are based on the motivational planning rules developed by del Soldato (1994), as we will see in section 7.4.2.
    - **Affective Dialogue rules**: these are also partially based on the rules developed by del Soldato (1994). These rules are responsible for the selection of appropriate affective feedback, given the student's performance for the current lesson.
    - **Motivation diagnosis rules**: these are the rules validated in chapter 6, and together with the Affective Dialogue rules, help us to infer the motivational state of the student, given his performance and his response to the feedback provided.
    - **Combination rules**: given that we have two models of the student's motivational state (his own self-reported model and the system inferred model), we need to combine these two models into one. This is the purpose of the combination rules.

3. The interaction history, where information about lessons already studied and the student's past performance is kept.

It is important to note that, although the main emphasis of this dissertation is on motivation diagnosis, and the evaluation in chapter 9 only deals with this issue, the prototype presented in this chapter also includes those components needed for motivational planning. Although these components were not strictly necessary for the evaluation of the motivation diagnosis capabilities in MOODS, they were developed as an attempt at completeness, since we wanted to develop a system that would simulate the whole tutoring process, although with the emphasis on motivation diagnosis.

Therefore, the motivational planning rules developed by del Soldato (1994) were analysed and transformed into a pattern that could be easily integrated with the motivation diagnosis rules that we had elicited thanks to the empirical studies discussed in previous chapters. This integration is not always complete, since the motivational planning rules developed by del Soldato (1994) were based on a simpler model of student's motivation. Nevertheless, it shows that the integration is possible and how a complete Affective Tutor could be developed.

In the following sections we give details of each of the components of MOODS.

7.3 Student models

7.3.1 Traits model

The traits model represent certain non-transient characteristics of the student, as seen in chapter 3. Four traits are represented in the model: *Control*, *Challenge*, *Independence* and *Fantasy* (see table 3.1 for the definitions of each trait). Each trait is represented as a value in a five point scale: *Very low*, *Low*, *Average*, *High* and *Very High*.

7.3.2 Reported, Inferred and Combined motivation models

The three motivational state models of the student (Reported, Inferred and Combined) represent transient characteristics of the student, which are given by five different factors: *Satisfaction*, *Confidence*, *Effort*, *Cognitive Interest* and *Sensory Interest*. As per the trait characteristics, each of the motivational factors is represented as a value in a five point scale: *Very low*, *Low*, *Average*, *High* and *Very High*.

The reported model is based on the self-report method seen in chapter 4, and it represents the student's own report of his motivational state. As was discussed in chapter 4, the use of self-report as a way of detecting the student's motivational state has a number of limitations, mainly:

- that the student can fail to update the self-reported model due to various reasons, for example: the task is too engaging; the task is too boring; updating the model does not seem to have any effect on the instruction; etc.

- that the student might be tempted to "please" the instructional system by providing artificially positive readings of his motivational state.

Thus, an Affective Tutor should not rely exclusively on self-report for the detection of the student's motivational state. For this reason we performed the Motivation Detection Study and the Validation Study described in chapters 5 and 6, in order to elicit a set of motivation detection rules that would enable an Affective Tutor to infer the motivational state of the student based on his interaction with the system. This inferred motivational state is represented by the same variables as the self-reported model, but it is stored as a separate model.

Finally, we need to combine these two sources of information about the student's motivational state into one, as the information in one of the models can be incomplete or even contradictory to the other model. This is the purpose of the *Combined model*, which is also represented by the same variables as the *Self-reported* and the *Inferred* models, but it is generated by applying the *Combination Rules* to these latter models. This is the model that it is considered to have the most accurate picture of the student's motivational state, and as such it is the model used as the source of information in the Motivational Planning rules.

7.4 Knowledge-based rules

7.4.1 Incorporating the instructional planning rules from del Soldato (1994)

As mentioned earlier, the motivational planning rules (see section 7.4.2) and the affective dialogue rules (see section 7.4.3) implemented in MOODS are partially based on the instructional planning rules implemented by del Soldato (1994). In this section we explain how we reused the rules developed in her thesis work in order to adapt them to the design of MOODS.

As explained in section 2.5.3, the instructional planning rules developed by del Soldato (1994) are composed of a *Domain-based planner*, a *Motivational Planner* and a *Negotiation Planner*[1]. In order to make the implementation of MOODS simpler, we subsumed these three planners into one. As a result of this process we ended up with a number of rules that represent quite accurately the rules developed by del Soldato (1994), but without the complexity of three different planners. The resulting rules can be seen in table 7.1.

From the three planners developed by del Soldato (1994), it is the *Negotiation Planner* the one that has the biggest importance. This planner "negotiates" between the *Domain-based* planner and the *Motivational Planner* in order to decide which actions to take. Therefore, in order to combine the three planners we started by looking at each of the rules in the *Negotiation Planner* and working "backwards" to find under which conditions each of these rules would be applied.

Let's illustrate this by presenting in detail the process followed for the rule *Sol1* in table 7.1. The last column in table 7.1 represents the rule from the work by del Soldato (1994) on which the given rule is based. Thus, we can see that the rule *Sol1* was derived from rule *N1*.

In table 7.2 we present all the rules from the *Domain-based planner*, the *Motivational Planner* and the *Negotiation planner* that were used in order to derive the rule *Sol1*.

Let's start by looking at the preconditions of *N1* in table 7.2(c). These are:

- The *Domain-based planner* suggested a problem of a harder type.
- The *Motivational-planner* suggested:
  - to increase the experience of success
  - not to stimulate challenge

In figure 7.2 we present a schematic representation of how the preconditions and actions of all the rules used in deriving the rule *Sol1* relate to one another. In this figure we can see the preconditions of *N1* under the Preconditions section of rule *N1*.

[1]These rules can be found in (del Soldato, 1994) in tables 4.10, 4.11 and 4.12 in pages 42, 43 and 46, respectively.

Rule	Quality	Give up	Help requested	Number of give ups	Past Performance	Confidence	Effort	Independence	next(Difficulty)	Assessment	Comment	Help	Based on
	PERFORMANCE			*INTERACTION HIS.*		*MOTIVATION MODEL*			*OUTPUTS*				
Sol1	≥Ave					< Ave	≥ Ave		=	Right			N1
Sol2	≥Ave					≥ Ave	< Ave		++	Right	Challenge		N2
Sol3	≥Ave					< Ave	< Ave		+	Right	Promotion		N3
Sol4	≥Ave						> Ave		+	Right	Praise-perf		N6
Sol5	≥Ave								+	Right			D1
Sol6	<Ave				≥ high	< Ave	< Ave		=	Wrong	Prev-success		N10
Sol7	<Ave					≥ Ave	< Ave					Surprise result	N5
Sol8	<Ave					< Ave						Next step	N4
Sol9	<Ave						> Ave		=	Wrong	Praise-eff		N7
Sol10	<Ave								=	Wrong			D2
Sol11		Yes		≤ Ave	≥ high	< Ave	< Ave		=			Prev-success	N10
Sol12		Yes		≤ Ave		≥ Ave	< Ave				Trying harder	Next step	N9
Sol13		Yes		≤ Ave		< Ave						Next step	N8
Sol14		Yes							=				D3
Sol15			Yes			≥ Ave		< Ave			Enc. indep.	Skip help	N11, N12
Sol16			Yes			< Ave						Present (specific)	N13
Sol17			Yes									Present step	D5

Table 7.1: Instructional Planning rules adapted from del Soldato (1994)

Rule	Student model / History	Action
D1	*problem-state* = *succeeded*	provide *assessment type right* suggest *problem type harder*

(a) Domain-based planner

Rule	Student model / History	Top-Level-Tactics	Tactic
M1	*conf-value* < *conf-threshold*	–	*increase confidence*
M2	*effort-value* < *medium*	–	*increase effort*
M7	*problem-state* = *suceeded*	*increase confidence*	*increase experience success*
M11	*problem-state* = *succeeded*	*increase effort*	*stimulate challenge*

(b) Motivational planner

	DOMAIN-BASED PLAN	MOTIVATIONAL PLAN	NEGOTIATION PLANNER	
Rule	Action	Tactic	Delete action	Add action
N1	suggest *problem type harder*	increase *experience success* not *stimulate challenge*	suggest *problem type harder*	suggest *problem type similar*

(c) Negotiation planner

Table 7.2: Rules from del Soldato (1994) used to derive rule *Sol1*.

At the same time, an arrow links each of these to the rule that satisfies each precondition. For example, we can see that rule *M7* satisfies the precondition *increase experience success*. And this is so when the *problem-state* = *succeeded* (i.e. when the student succeeded in performing the given task) and the tactic *increase confidence* exists. *Problem-state* = *succeeded* does not have any further preconditions, which we represent in figure 7.2 by printing it in bold. But rule *M7* also has as one of its preconditions that the tactic *increase confidence* has been generated, which is satisfied by the rule *M1*, when *conf-value* < *conf-threshold*.

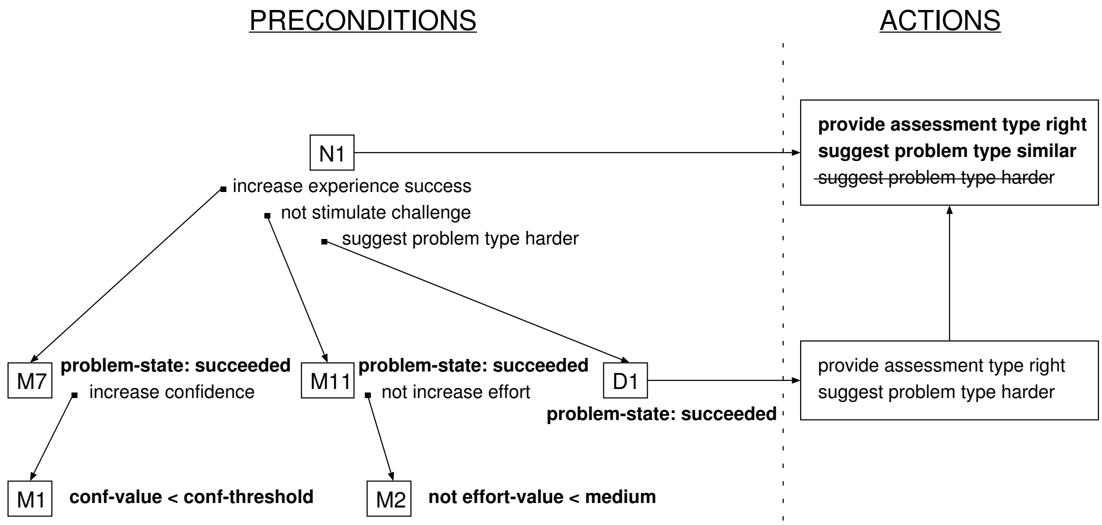


Figure 7.2: Relationship between preconditions and actions of the rules used to derive *Sol1*.

By proceeding similarly for all the other rules in figure 7.2, we find that the actual preconditions to apply for the rule *N1* are:

- problem-state: succeeded
- conf-value < conf-threshold
- not (effort-value < medium)

These are the preconditions for rule *Sol1* in table 7.1. If we now look at the actions to be generated by applying rule *Sol1*, we can see in table 7.2 that only two of the rules generate actions: *D1* and *N1* (the other rules generate tactics that can be further used by other rules, but not actions).

Rule *D1* generates two actions (which can be seen in figure 7.2 in the box to the right of the rule *D1*):

- provide assessment type right
- suggest problem type harder

But when the appropriate preconditions exist and the rule *N1* applies, it modifies the actions generated by *D1*, deleting the action *suggest problem type harder* and adding the action *suggest problem type similar*. Thus, the rule *Sol1*, which subsumes all the rules represented in figure 7.2 applies when the preconditions indicated above are satisfied, and as a result of applying the rule *Sol1* two actions are taken:

- provide assessment type right
- suggest problem type similar

This can be seen in table 7.1, although the terminology in that table has been changed slightly in order to better accommodate it to the design of the knowledge-based rules implemented in MOODS.

By applying a similar process to all the rules in the different planners developed by del Soldato (1994) we collated all the rules in table 7.1, which are the basis for the rules for Motivational Planning and Affective Dialogues in MOODS, which are presented in sections 7.4.2 and 7.4.3 respectively. The different values and categories of the rules are also explained in those sections.

7.4.2 Motivational planning rules

As mentioned in section 7.4.1, the motivational planning rules in MOODS are based on the instructional rules developed by del Soldato (1994), although the format and some of the details of those rules had to be changed in order to accommodate them to the design of MOODS. The resulting motivational planning rules implemented in

MOODS can be seen in table 7.3, where the last column indicates which rule from table 7.1 is the origin of the given rule.

For example, rule *MP1* is based on rule *Sol1*. The only difference is that the rule *MP1* lacks three of the *Output* columns, namely: *Assessment*, *Comment* and *Help*. These columns refer to the type of feedback to provide to the student, and as such are dealt with in MOODS by the *Affective Dialogue rules* in section 7.4.3.

In order to make table 7.3 easier to understand, we named its input factors by the same naming conventions as used for the rules elicited via the Motivation Diagnosis Study described in chapter 5. These are[2]:

- Performance
  - **Quality**. Correctness of the answers provided to the exercises.
  - **Give up**. Whether the student chose to give up the lesson or not.
- Interaction History
  - **Number of give ups**. Number of times the student has attempted to give up the present lesson.
  - **Past Performance**. Average correctness of previous exercises.
- Motivation model
  - **Confidence**, **Effort**. Value of the corresponding factor in student's motivational model.

As per the values of the individual rules, *Ave* represents the median in our five points scale. Thus, rule *MP1* indicates that

if

1. the *Quality* of the exercise performed was at least average (i.e. *Average*, *High* or *Very High*),

[2]These definitions (except for *Number of give ups* and *Past Performance*) were given in page 101, and are reproduced here for convenience. We do not include here the definitions for the factors that do not appear in tables 7.3 or 7.4.

Rule	Quality	Give up	Number of give ups	Past Performance	Confidence	Effort	next(Difficulty)	Based on
	PERFORMANCE		*INTERACTION HIS.*		*MOTIVATION MODEL*		*OUTPUTS*	
MP1	$\geq$Ave				$<$ Ave	$\geq$ Ave	$=$	Sol1
MP2	$\geq$Ave				$\geq$ Ave	$<$ Ave	$++$	Sol2
MP3	$\geq$Ave				$<$ Ave	$<$ Ave	$+$	Sol3
MP4	$\geq$Ave					$>$ Ave	$+$	Sol4
MP5	$\geq$Ave						$+$	Sol5
MP6	$<$Ave			$\geq$ high	$<$ Ave	$<$ Ave	$=$	Sol6
MP7	$<$Ave					$>$ Ave	$=$	Sol9
MP8	$<$Ave						$=$	Sol10
MP9		Yes	$\leq$ Ave	$\geq$ high	$<$ Ave	$<$ Ave	$=$	Sol11
MP10		Yes					$=$	Sol14

Table 7.3: Motivational Planning rules.

2. the student's confidence is lower than average (i.e. *Low* or *Very Low*),

3. and the student's effort has been at least average

then

- the next exercise we will present to the student will be of the same difficulty as the present one.

As we can see, the Motivational Planning rules implemented in MOODS are few and lack the complexity of the Motivation Diagnosis rules, as seen in chapter 6. This is the case since we implemented in MOODS a variation of the rules developed by del Soldato (1994), which are based on a simpler student model.

Nevertheless, these rules are sufficient for the development of MOODS. Although it would be desirable to develop a fuller set of Motivational Planning rules for an actual *Affective Tutor*, the main goal of MOODS was to develop an *Affective Tutor Simulation* in which the Motivation Diagnosis techniques explored in this dissertation could be evaluated. As such, the importance of the Motivational Planning rules in MOODS is smaller, although they were necessary in order to develop a believable simulation of a tutoring system.

7.4.3 Affective Dialogue rules

An "ideal" ITS should use language to communicate with its student. Although particular instructional units can be ingeniously designed to avoid language, a complete instructional interaction requires a number of tasks for which it would be very difficult not to use language. This is so not only in the arts and humanities. As du Boulay and Luckin (1999) comment, "[...] even in mathematics and science, despite all the 'apparatus' of representations and formal manipulations on those representations, an essential part of the 'glue' which binds what otherwise might be a fragmented understanding are conversations both with others and with ourselves."

At the same time, verbal communication can be used as a further means to motivation diagnosis in an Affective Tutor, as it offers some benefits over other types of motivation diagnosis:

- Once a verbal interaction or dialogue is started, the student is somehow 'forced' to answer the questions posed by the system. This avoids a common problem of the self-report method (see chapter 4); namely, that lack of student interaction with the self-report facilities can be misinterpreted as a constant value for the variables represented in the self-report interface.

- Through verbal interaction it is possible to infer certain aspects of a student's motivational state, while it might appear to him that the goal of the dialogue is somehow different. For example, asking the student if he would like to continue with a lesson of similar difficulty could inform us about his confidence, but without asking directly about it. People tend to treat machines as other social beings (Reeves and Nass, 1998), which can lead to the student 'lying' about his motivational state if asked directly, in order to 'please' the ITS. By obscuring the actual purpose of the verbal interaction we might be able to avoid this, and therefore be able to obtain more accurate information about the student's motivational state.

In this section we present the Affective Dialogue rules implemented in MOODS. This set of rules has been partially based on the work by del Soldato (1994), but we did not intend to provide a complete or comprehensive solution for generating Affective Dialogues. The rules implemented in MOODS are very basic[3], and their only purpose is to illustrate the use that language could have in detecting a student's motivational state.

Providing feedback

Given the simulated performance of a student, and in order to decide which feedback to provide, MOODS makes use of the rules presented in table 7.4. As explained in section 7.4.1, these rules are based on the instructional rules developed by del Soldato (1994). As per the motivational planning rules, the format and some of the details of

[3]More interesting approaches to generating affective language exist. See for example (de Rosis and Grasso, 1999; de Rosis *et al.*, 1999) for an approach in which natural language generation techniques are used for affective text generation, and (Porayska-Pomsta *et al.*, 2000) for more linguistically motivated research on generating teachers' language in educational dialogues.

those rules had to be changed. The last column of table 7.4 indicates the rule from table 7.1 on which the given rule is based.

For example, rule *AF1* is based on rule *Sol1*. The only difference between these rules is that the rule *AF1* lacks the column *next(Difficulty)*. That column refers to the next exercise to select, and this issue is dealt with in MOODS by the *Motivational Planning rules* as seen in section 7.4.2.

These rules are very similar to the *Motivational Planning* rules, but the output of these rules refers to the feedback to provide to the student, which is divided into two parts: a) *Assessment* and b) *Comment*.

The assessment refers to the quality of the student's performance, while the comment is an elaboration on the assessment. For example, rule *AF2* has as its outputs:

- **Assessment**: Right
- **Comment**: Challenge

Thus, under the conditions given by rule *AF2*, MOODS would provide an assessment of type *Right*, and would provide a comment of type *Challenge*. We can see an example of the realization of this type of feedback in table 7.5. The first two columns in this table refer to providing feedback, while the last two columns refer to the detection of the student's motivational state, which is dealt with in the next section.

Thus, for an assessment of type *Right* and a comment of type *Challenge* we would provide the following feedback: *"That was very good. Let's try something much harder now!"* (As mentioned in table 7.4, the assessment of type *right* can have three different values, according to the value of the *Quality of the performance*. This is reflected in table 7.5 by providing three different choices between brackets).

Detection of student's motivational state

As mentioned earlier, the replies that the student gives to the feedback can be used to infer his motivational state. Without attempting to attach any validity to them, we present in table 7.5 some rules that could be used in MOODS to diagnose the motivational state of the student given the reply of the student to the feedback provided.

Rule	Quality	Give up	Number of give ups	Past Performance	Confidence	Effort	Assessment	Comment	Based on
	PERFORMANCE		*INTERACTION HIS.*		*MOTIVATION MODEL*		*OUTPUTS*		
AF1	$\geq$Ave				$<$ Ave	$\geq$ Ave	Right[a]		Sol1
AF2	$\geq$Ave				$\geq$ Ave	$<$ Ave	Right	Challenge	Sol2
AF3	$\geq$Ave				$<$ Ave	$<$ Ave	Right	Promotion	Sol3
AF4	$\geq$Ave					$>$ Ave	Right	Praise-perf	Sol4
AF5	$\geq$Ave						Right		Sol5
AF6	$<$Ave			$\geq$ high	$<$ Ave	$<$ Ave	Wrong[b]	Prev-success	Sol6
AF7	$<$Ave					$>$ Ave	Wrong	Praise-eff	Sol9
AF8	$<$Ave						Wrong		Sol10
AF9		Yes	$\leq$ Ave	$\geq$ high	$<$ Ave	$<$ Ave		Prev-success	Sol11

Table 7.4: Affective Dialogue (AFDI) rules.

[a] As per the values to quality, there are 3 values for a positive assessment, depending on whether the quality was Average, High or Very High.

[b] Similar to the positive assesment values, there are 2 values for a negative assessment, corresponding to the Low and Very Low quality values.

Asessment + Comment	Sample	Student	Diagnosis
Right	That was [OK / very good / excellent!]	Thanks, it was easy.	*Effort*: Dec
		Thanks, but it was hard!	*Effort*: Inc
Right + Challenge	That was [OK / very good / excellent!] Let's try something much harder now!	Yes, I think I can do it.	*Confidence*: Inc
		Perhaps not much harder.	*Confidence*: Dec
Right + Promotion	That was [OK / very good / excellent!] Let's try now a more difficult exercise.	Ok, I will try.	*Confidence*: =
		I don't think I will manage.	*Confidence*: Dec
Right + Praise-perf	That was [OK / very good / excellent!] You did very well.	Thanks, but it was a bit difficult.	*Confidence*: Dec
		Thanks, it was very enjoyable.	*Satisfaction*: High
Wrong	[Sorry, you did poorly. / You did very poorly!]	But it was so difficult!	*Effort*: Inc, *Confidence*: Dec
		But it was so boring!	*Satisfaction*: Dec
Wrong + Prev-success	[Sorry, you did poorly. / You did very poorly!] You have done very well previously.	OK, I will try a bit harder.	*Effort*: Inc
		Well, probably I was just lucky before.	*Confidence*: Dec
Wrong + Praise-eff	[Sorry, you did poorly. / You did very poorly!] But it was a good effort.	Thanks, but it was hard!	*Effort*: Inc
		I don't think I can solve this.	*Confidence*: Dec
Prev-success	You have done very well previously.	OK, I will try a bit harder.	*Effort*: Inc
		Well, probably I was just lucky before.	*Confidence*: Dec

Table 7.5: Surface realization of AFDI rules and the effect of a student's reply on his motivational state.

These rules have not been validated, but we present them here, as they provide examples of the types of rules that could be incorporated into a more sophisticated tutoring system, and that would help to make predictions about the motivational state of the student based on an Affective Dialogue.

As we can see in table 7.5, every type of feedback can have a number of possible replies by the student. Depending on the selection of the student, we could sometimes infer some changes in his motivational state. Thus, when the feedback provided is of type *Right*, for example *That was excellent!*, the student could choose from two possible replies:

1. Thanks, it was easy.
2. Thanks, but it was hard!

Each of these replies conveys a different message about the effort required by the student to complete the task. A reply of *Thanks, it was easy* could mean that the student did not have to put much effort into the task, and therefore we would decrease the *Effort* value of his motivational model. On the other hand, a reply of *Thanks, but it was hard!* could mean that it was a difficult task for the student, and therefore we would increase the *Effort* value of his motivational model.

As mentioned earlier, these rules are very basic and have not been validated, but they help to illustrate the role that language could have in an Affective Tutor.

7.4.4 Motivation diagnosis rules

The most substantial part of knowledge in MOODS is given by the Motivation Diagnosis rules, which were elicited via the study presented in chapter 5 and validated via the study presented in chapter 6.

These rules were presented and discussed earlier on, so we do not include them here again. All the Motivation Diagnosis rules implemented in MOODS are those named in table 6.4 on page 116. The actual details of all the rules can be seen in tables 5.4 to 5.13, on pages 94 to 100. We just reproduce here one of these tables for reference purposes (see table 7.6).

Rule	Quality	Quantity	Speed	pre(Give up)	Give up	Hesitation	Performed in order	Difficulty	Output	Based on
	PERFORMANCE							*TE. MAT.*		
IE1					No	High			Inc	3a
IE2		Very high							High	5a
IE3		High			No				High	5b
IE4							Yes		Inc	4g
IE5				Yes					High	4o
IE6		High	Avg					High	High	6d
IE7	High		Slow						High	10h

Table 7.6: Increase Effort diagnosis rules (reproduced from table 5.8).

7.4.5 Combination rules

The issue of combining the *self-reported* and the *inferred* models is a very complex issue, and therefore we provide here only a very coarse approach to it, which is based on two rules:

1. If a motivational factor in either the *self-report* model or the *inferred* model has been updated but not in both, then we consider the updated value to be the correct one, and the value of the factor in the *combined* model will be updated accordingly.

2. If a motivational factor in both the *self-report* model and the *inferred* model has been updated, then we consider the average of these values to be the correct one, and the value of the factor in the *combined* model will be updated accordingly.

This approach to combining the *self-reported* and the *inferred* models is only a first approximation, and ideally a further study should be conducted to understand how best we should approach this issue.

7.5 Interaction history

In its present state, the Interaction History in MOODS is based on the information required by the implemented rules presented in the previous sections.

As introduced in the previous sections, we model six variables: *Past Performance*, *pre(quality)*, *pre(quantity)*, *pre(speed)*, *pre(giveup)*, and *pre(diff)*.

The variable *Past Performance* is used by some of the rules of the Motivational Planning and the Affective Dialogue rules. It is an indication of the quality of the past performance of the student.

The remaining variables are used by the Motivation Diagnosis rules, and refer to qualities of the previous exercise. *pre(diff)* refers to the difficulty of the previous exercise. The other variables refer to the performance of the student during the previous exercise, and they were explained in page 101 of chapter 5.

7.6 Discussion

In this chapter we have presented the design of MOODS, a shell system which incorporates all the motivational planning and motivation diagnosis knowledge described in this dissertation.

By developing MOODS, we created a tool that allows us to easily simulate a possible interaction with a student, based on the knowledge presented in this chapter. A sample interaction with MOODS is given in chapter 8.

By using MOODS in this fashion, we prepared a number of instructional interaction simulations in order to evaluate the approach to motivation diagnosis presented in this dissertation. The description of this evaluation is given in chapter 9.

Chapter 8

Example interaction

8.1 Introduction

As explained in chapter 7, MOODS was developed as a shell where all the knowledge about motivation detection presented in previous chapters could be easily tested. In this chapter we introduce MOODS and give an example of how an instructional interaction can be simulated. We start by giving a detailed account of the MOODS interface in section 8.2.

It is important to note that MOODS was developed as a tool for our own use, which would allow us to easily simulate instructional interactions. Thus, the interface presented in this chapter was mainly meant as a tool to help us develop the evaluation presented in chapter 9, and therefore not a great attention was given to its design. Nevertheless, an improved version of MOODS could be an interesting tool in its own right, as noted in section 10.3. For example, MOODS could be used as a teacher training tool, in which a teacher could simulate an instructional action in order to see the influence that this action could have on the motivational state of a student.

8.2 MOODS Interface

In figure 8.1 we can see the initial view of the MOODS interface before the start of the interaction. The interface is divided into three main areas:

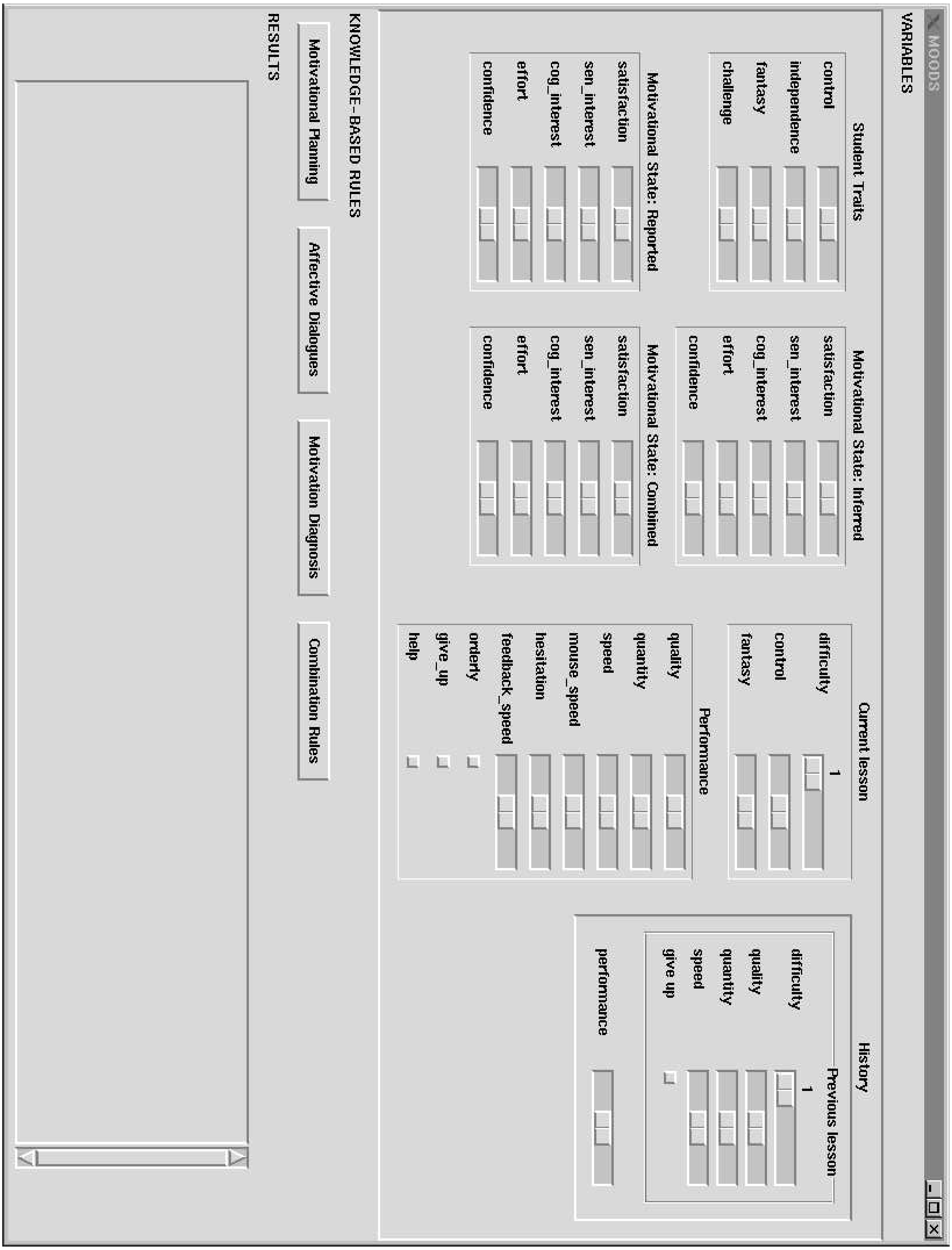


Figure 8.1: MOODS Interface

1. **Variables** frame. In this frame we have a representation of all the variables used by the MOODS knowledge-based rules. These variables represent the student's models, the characteristics of the teaching materials and the details of the student's simulated performance. More details about the variables frame are given in section 8.2.1.

2. **Knowledge-based rules** frame. This frame contains four buttons, one for each type of knowledge-based rules developed in MOODS: *Motivational Planning*, *Affective Dialogues*, *Motivation Diagnosis* and *Combination Rules*. We can press any of these buttons in order to see the result of applying the corresponding rules in the given simulated situation. That is, given the instructional situation represented in the Variables frame, we would press any of these buttons (for example Motivational Planning) in order to obtain the inferences made by those rules.

3. **Results** frame. It is in this frame that the inferences made by the knowledge-based rules are presented. Examples of these are given in section 8.3.

8.2.1 Variables frame

As mentioned above, the variables frame represent a number of characteristics of the simulated instructional interaction. It is divided into three main sections. From left to right these are:

1. **Student models section.** This section is formed by the four frames to the left, representing the student *traits model* and the three student motivational state models: *reported model*, *inferred model* and *combined model*.

2. **Current lesson section.** This section is formed by two frames:

   (a) The top frame represents the characteristics of the currently simulated lesson, in terms of its *difficulty*, and the level of *control* and *fantasy* that it offers.

(b) The bottom frame represents the student's performance on this lesson, represented by the factors introduced in chapter 5, such as *quality*, *quantity*, *speed*, etc.

3. **History section.** This section is given by:

   (a) Characteristics relating to the previous lesson: its *difficulty*, and a number of characteristics of the student's performance (such as *quality*, *quantity*, etc.).

   (b) A representation of the student's overall *performance* quality for all the past lessons.

All the variables in the *Current lesson section* and the *History section* represent the input variables to the knowledge-based rules described in chapter 7. Thus, by modifying these values we can simulate different instructional situations, an example of which is shown in section 8.3.

These variables can be divided into three groups according to the possible values that they can take:

1. Most of the variables are given by a five value scale, which represent the values *Very low*, *Low*, *Average*, *High* and *Very High*[1]

2. Some other variables are boolean variables, that is, variables that can take a Yes or No value. These are *Orderly*, *Give up* and *Help*, and they are represented by check buttons.

3. The variables representing the difficulty of the current and the previous lesson are represented as a scale with possible values 1 to 5.

8.3 Instructional Simulation example

Let's imagine we want to simulate an instructional situation in which:

[1]This is represented internally as: -10, -5, 0, 5, 10. In the remaining of the chapter we use this scale and not the equivalent categorical scale.

- The student performed the current lesson very well.
- The student completed most of the exercises in the lesson.
- The student showed little hesitation.
- The student didn't ask for help.
- The student reported that he had to put a lot of effort in doing the exercises.

We represent this situation in MOODS by updating the five variables highlighted in the Variables frame in figure 8.2. Ideally we would represent every instructional situation in terms of all the variables in the Variables frame. Nevertheless, if we want to represent a situation in terms of only some of the variables, we update only those variables in MOODS, and set the remaining variables to their average value, as is shown in figure 8.2.

Once an instructional situation has been represented by updating the corresponding variables, we obtain the resulting inferences by pressing the buttons in the Knowledge-based rules frame. In figure 8.2 we can see the results of pressing the four buttons in the order corresponding to the numbers attached[2].

1. We pressed first the button *Combination Rules*, to get the results of combining both the inferred and the reported models. The corresponding inference can be seen in the Results frame in the first position. As we can see, only the reported value of the effort variable has changed, and as such the combined value should be the same as the reported one. We update the effort variable in the Combined Motivational model, and continue with the rest of the inferences.

2. We press the *Affective Dialogues* button to see the Feedback that MOODS suggests. As we can see in the Results frame, given the present instructional situation, MOODS suggests to provide a **Very good** assessment, and a *Praise Per-*

[2]In the results frame, the latest result is always presented on the top of the frame.

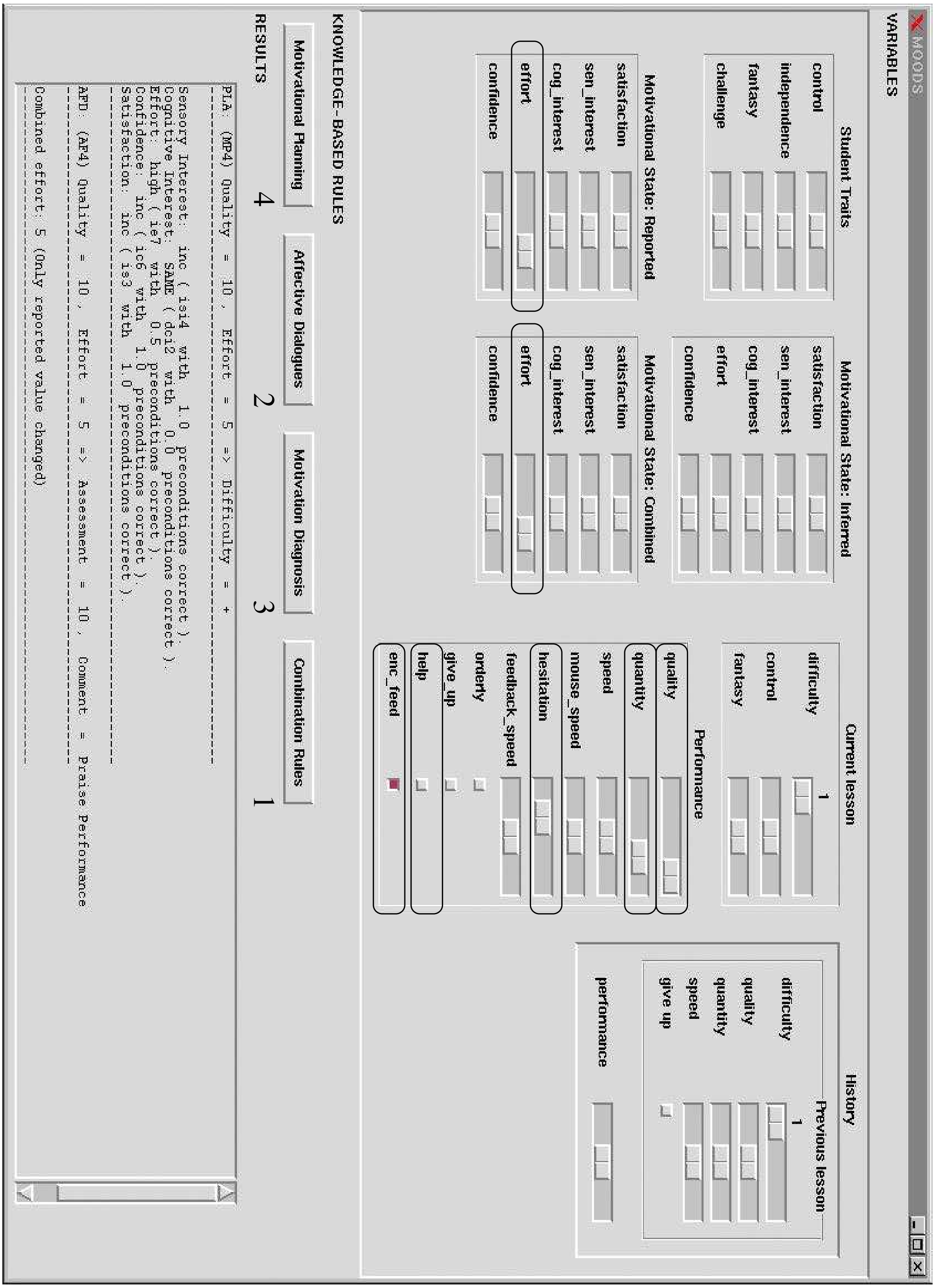


Figure 8.2: Sample interaction with MOODS

formance comment. We see between brackets that this is due to rule AF4 being fired, because *Quality = 10* and *Effort = 5*.

3. We press the *Motivation Diagnosis* button to see the changes to the Inferred Motivational Model that MOODS suggests. As we can see in the Results frame, MOODS suggests to leave the *Cognitive Interest* value with the same value, and to set the values for *Effort*, *Confidence* and *Satisfaction* as High. Details about how these are calculated are given in section 8.3.1.

4. Lastly, we press the *Motivational Planning* button to see the suggestion of the difficulty of the next exercise to be presented, given the present instructional interaction and motivational state of the student. In the example given in figure 8.2 we can see that an exercise of greater difficulty is presented by applying rule MP4, as *Quality = 10* and *Effort = 5*.

8.3.1 Motivation diagnosis inferences

As we saw in chapter 6, the rules implemented in MOODS are many, and therefore conflicts can occur when we try to match them against a simulated instructional situation. At the same time, the implemented rules don't cover all possible instructional situations, and therefore under many conditions we would obtain no motivation diagnosis inferences.

In order to avoid this, we implemented the motivation diagnosis rules in MOODS in such a way that for each motivational factor, the rule for which most of its preconditions are met will be the rule that will indicate the inference to make. That is, for each of the motivational factors (satisfaction, effort, etc.) we consider all the possible rules and select the one that has the greatest percentage of valid preconditions. If at least 50% of its preconditions are met then this rule would indicate how the corresponding factor should be changed. Otherwise, we consider that there is not sufficient evidence for making an inference for this factor, and would leave the value of the corresponding factor unchanged.

Thus, in figure 8.2 we see that rule IS3 was chosen to infer the changes to make to the Satisfaction factor. In table 8.1 we can see in bold the preconditions met for

Rule	Quality	Give up	pre(Difficulty)	Difficulty	Fantasy	Feedback	Confidence	Effort	Cognitive interest	Challenge	Output
	PERFORM.		*TEACH. MATERIALS*				*MOT. MODEL*			*TRAITS*	
IS1	**High**						High		High		High
IS3	**High**					**Enc**					Inc
IS4	**High**					**Enc**			High		High
IS6	**High**		X	>X				**High**		High	High
DS1	Low						Low	Low			Low
DS2	Avg	Yes									Low
DS6	V. low				V. low						V. low

Table 8.1: Number of preconditions met for some of the Satisfaction detection rules.

some of the satisfaction diagnosis rules given the situation presented in figure 8.2. As we can see, no preconditions are met for any of the rules that infer a low value for satisfaction. On the other hand, there is a varying number of preconditions met for those rules that infer a high value for the satisfaction motivational factor. From these, IS3 has the largest percentage of preconditions met: all of them. Therefore, we would consider rule IS3 to be the most appropriate rule to apply under the given situation.

By following a similar procedure for all the motivational factors, we can infer how to update the complete inferred motivational state model, as seen in figure 8.2.

8.4 Discussion

In this chapter we have presented very briefly an example interaction with MOODS. As we have seen, MOODS is a simple tool which incorporates all the motivational planning and motivation detection knowledge described in previous chapters of this dissertation. At the same time MOODS allows us to easily test the motivation diagnosis knowledge implemented in it by simulating an instructional interaction in terms of a number of instructional variables, as described above.

The evaluation of the motivation diagnosis knowledge implemented in MOODS is left for chapter 9. In that chapter we compare the inferences made by human teachers to the inferences made by MOODS, under a number of instructional situations. By following a similar procedure as that described in this chapter, we gathered information about the inferences made by MOODS under a number of instructional situations, which we could then compare against the inferences made by human teachers under the same situations.

Chapter 9

Evaluation

9.1 Introduction and goal of evaluation

In chapters 7 and 8 we have presented MOODS, the simulation system that we implemented, and which incorporates all the motivation detection knowledge discussed in this dissertation.

Evaluation of the approach to motivation diagnosis implemented in MOODS is a difficult issue, since motivation is in itself a psychological construct which is very difficult to measure. What can be done is to compare the similarities or otherwise of the motivational values inferred by MOODS with the values inferred by human teachers in similar situations.

In this chapter we present the evaluation of MOODS in the above mentioned terms. Although MOODS was implemented as an actual working system, we decided to evaluate it through an on-line study similar to the validity study described in chapter 6. The major advantage of this approach is the large number of participants that could be obtained. During this evaluation study participants were presented with an instructional interaction and were asked to infer the motivational state of a hypothetical student. Comparison with the values inferred by MOODS gave us a chance to evaluate the motivation detection approach implemented in MOODS.

9.2 Materials

As mentioned above, the evaluation study took the form of an on-line questionnaire. Participants were asked to imagine a situation in which a hypothetical student is using a computerised instructional system. For this, they were presented with a number of instructional settings (given as a number of simple statements), an example of which can be seen in figure 9.1.

The instructional setting was given as a number of simple sentences, representing a MOODS state before the motivational state of the student was modified.

Two different simulations of MOODS were developed as questionnaires, each of them consisting of six instructional settings. Obviously, these two simulations of instructional interactions do not represent all the possible avenues that a MOODS simulation can take. Nevertheless, they provide us a reasonable sample of instructional situations for which we can compare the motivational state values inferred by the human teachers and the MOODS system. Details of these two different simulations are given in section 9.2.1. As we can see in figure 9.1, a sample instructional setting was given as:

- He performed the lesson very well.
- He completed most of the exercises in the lesson.
- He showed little hesitation.
- He didn't ask for help.
- He reported that he put a lot of effort in doing the exercises.
- He was given the feedback: **That was excellent! You did very well.**

For each of the instructional settings, participants were asked two questions:

1. **What do you think his Motivational State will be at this point?** This question involved making a prediction about the motivational state of the student given the instructional setting presented. For each of the given factors the participant was

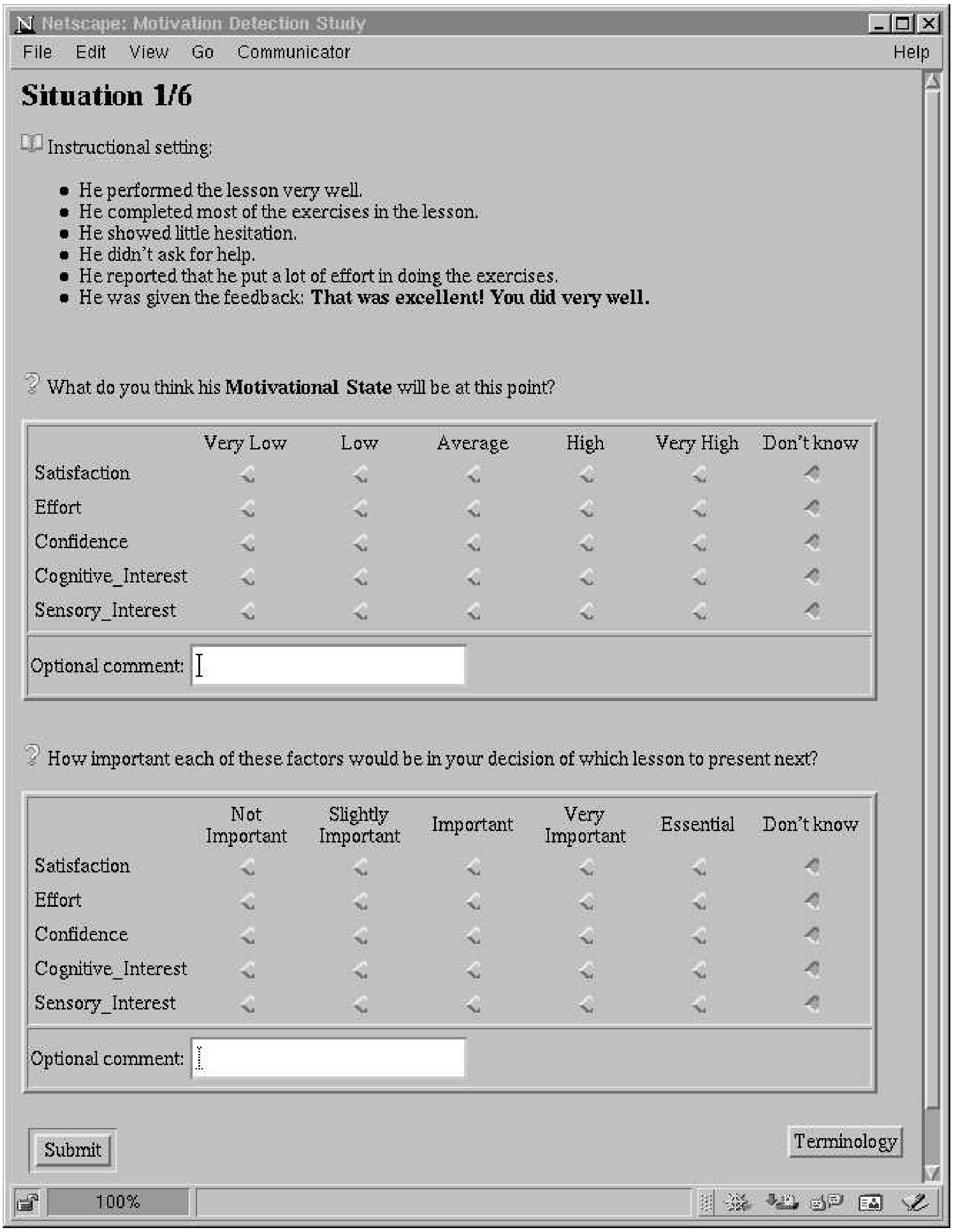


Figure 9.1: Evaluation Study sample question.

asked to choose the value that she thought was most appropriate, or to choose "Don't know" if she could not make a decision about that particular factor.

As possible options for each factor we gave them five values (besides the *Don't know* option), corresponding to the five possible values implemented in MOODS for each factor: *Very Low*, *Low*, *Average*, *High* and *Very High*.

2. **How important each of these factors would be in your decision of the next instructional step?** The second question referred to the importance that the participant would give to each of the motivational factors if she had to decide which instructional step to take next. That is, given the instructional setting presented and given her predictions for the student's Motivational State, she was asked to imagine that she was in control of the instruction. And the participant was asked to consider which of the motivational factors would she consider essential, not important, etc?

   The possible options for the second question were also five values (besides the *Don't know* option), namely: *Not important*, *Slightly important*, *Important*, *Very Important*, and *Essential*.

At the same time, the participants were encouraged to provide any optional comments for any of the questions if they wished to.

9.2.1 Simulations

As explained above, the instructional settings presented to each participant were organized into two different simulations. Each simulation consisted of six instructional settings, representing a likely interaction of a student with a tutoring system. Each of the simulations was developed by us, as an attempt to recreate a likely interaction of a student. Both simulations were developed with a prototypical student in mind, as we explain below.

Simulation 1. This simulation represents an interaction of a very confident student, with a high degree of control and fantasy, with an unresponsive tutoring system. We attempted to capture the situation in which a student with high confidence in his skills is using a tutoring system that is not able to provide him with sufficiently challenging situations, and as a result the student's interest in the interaction decreases, ending with a very poor performance and giving up.

Simulation 2. This simulation represents an interaction of a very able student, who enjoys a high degree of challenge, and who is using a tutoring system that is able to provide him a challenging instruction. Throughout the simulation, the student performs well, and excels when a challenging situation is given to him, putting a lot of effort in performing the task at hand.

Ideally, we would have represented each situation by providing values for all the variables represented in MOODS. But this was not feasible, as this would have meant to provide too much information for each instructional situation: at least nineteen pieces of information. During the pilot study we concluded that an appropriate number of items to describe an instructional situation would be between six and eight items of information.

Thus, we developed the simulations by representing each situation only with this number of items of information. On one hand, this compromise meant that the instructional settings would not be as detailed as we would have hoped for. On the other hand, by keeping the instructional descriptions simple we made sure that each of them would be simple enough for participants to understand and to imagine the given situation.

In any case, in order to create instructional situations that would be sufficiently informative for participants to make inferences about the possible motivational state of the hypothetical student, we attempted to have a similar distribution of the instructional variables in the situation descriptions of our study as in the motivational diagnosis rules in MOODS.

We started by calculating how often each instructional variable appeared in the motivational diagnosis rules in MOODS, which we can see in table 9.1. For example, in this table we can see that the variable "Quality" is used in 21 occasions in the motivation diagnosis rules in MOODS (see tables 5.4 to 5.13 for a description of all the rules).

"Speed" is mentioned in 15 occasions, while variables such as "Challenge", "Feedback speed", etc. are used only once. In table 9.1 we can also see this information given as percentages. Thus, we see that if we take into account all the motivational rules, the different instructional variables are used in total 137 times. The percentage corresponding to the variable "Quality" is 15.33% , while it is only 0.73% for the variable "Challenge".

	N	%		N	%
Quality	21	15.33%	pre(Quality)	3	2.19%
Speed	15	10.95%	Fantasy (Trait)	2	1.46%
Difficulty	15	10.95%	Satisfaction	2	1.46%
pre(Difficulty)	11	8.03%	Effort	2	1.46%
Quantity	11	8.03%	Performed in order	2	1.46%
Give up	10	7.30%	Mouse movements	2	1.46%
Hesitation	9	6.57%	Control	2	1.46%
Feedback	5	3.65%	Challenge	1	0.73%
Control (Trait)	4	2.92%	Feedback speed	1	0.73%
Fantasy	4	2.92%	Help	1	0.73%
Confidence	4	2.92%	Reset game	1	0.73%
pre(Give up)	4	2.92%	pre(Speed game)	1	0.73%
Cognitive Interest	3	2.19%	Speed game	1	0.73%

Table 9.1: Occurrence of instructional variables in motivational diagnosis rules in MOODS.

In order to create the instructional situations for the simulations, we made use of this information to resemble as much as possible the same distribution of variables for the simulations and for the motivational diagnosis rules in MOODS. Thus, we wanted to assure that 15.33% of the variables used to describe the instructional situations for both simulations would refer to "Quality", while only 0.73% of the variables would refer to "Speed game". To obtain the same distribution was not possible, since to do this we would have had to express the instructional situations based on mainly only three or four different variables, but nevertheless, we attempted to create instructional situations with a similar distribution of instructional variables as the motivational rules in MOODS, incorporating when necessary variables that were seldom used in the rules implemented in MOODS.

Once it was decided the type of variables to use for each instructional situation, the actual values were selected based on the description of the simulations, as described above. For example, we decided that the variables that should appear in the first situation of the Simulation 1 (as seen in figure 9.1) involved the factors: "Quality", "Quantity", "Hesitation", "Help", "Effort" and "Feedback". The actual values for these variables were chosen by us based on the description of the simulation. As we said above, the Simulation 1 represents an interaction of a very confident and able student, and therefore, the values for the given variables reflect this: he performed the lesson very well, he showed little hesitation, etc.

The only variables that were not invented by us were those dealing with feedback and difficulty of the lessons, which were based on the actual responses given by MOODS. Thus, under the given situation, MOODS offers a very positive feedback (which is translated in the instructional situation as "That was excellent! You did very well.") and suggests a more difficult exercise, which is reflected in the second instructional situation for Simulation 1 (see the description of the first three situations of Simulation 1 below).

The complete description of both simulations is given in section D.2 of appendix D. For illustration we reproduce here the description of the first three situations of Simulation 1.

First three situations of Simulation 1.

1

```
He performed the lesson very well.
He completed most of the exercises in the lesson.
He showed little hesitation.
He didn't ask for help.
He reported that he put a lot of effort in doing the exercises.
He was given the feedback: That was excellent! You did very well.
```

2

```
This time he was given a more difficult lesson.
He reported he was very confident that he could do this lesson well.
He performed the lesson well.
He completed all the exercises in the lesson.
He didn't complete the exercises in the given order.
He didn't give up the lesson.
He was given the feedback: That was very good.
```

3

```
We know that he is a person that enjoys a high degree of control.
This time he was given a more difficult lesson.
He completed about half of the exercises in the lesson.
Of those exercises that he completed, he performed the lesson satisfactorily.
He submitted the lesson quickly.
He didn't complete the exercises in the given order.
He was given the feedback: That was OK.
```

9.3 Participants

As per the study described in chapter 6, we asked for collaboration in a number of mailing lists dealing with technology and education in order to obtain participants. The prerequisite for participating in the study was to have at least one year of teaching experience and, as a result of this, 39 participants volunteered to take part in this study.

Participants were allocated alternatively to one of the simulations as they volunteered, in order to ensure an unbiased selection and a similar number of participants per simulation. Thus, 19 participants followed the first simulation, while the remaining 20 followed the second simulation.

In figure 9.2 we see the distribution of years of teaching experience amongst the participants for both simulations. As we can see, the majority of participants were very experienced teachers (32 of the 39 participants had 4 years of experience or more). The subject area was not a precondition for participation in the study, and therefore there was an ample range of subjects taught by the participants.

The complete list of all the subjects taught by the participants is very varied and includes among others: Primary Education, Economics, Biology, Ecology, Anthropology, History, Music, English, French, Geography, Education, Medicine, Engineering, Architecture, Mathematics, Physics, Computer Science, Philosophy, Psychology and General Science.

Given the large number of subjects taught by the participants, a statistical analysis of the relation between the subject taught and the teaching experience was not feasible, but an informal analysis of the data showed that there was not any clear correlation between them. For example, the subjects taught by the five less experienced participants were: Architecture, CAAD, Computer Information Systems, English, Computer Technology, Psychology, Writing. The subjects taught by the five most experienced participants were: Primary Education, Anthropology, Linguistics, English, German, History, Metacognitive Strategies, Mathematics, Ecology and Computing.

9.4 Methodology

The evaluation was performed as an on-line questionnaire. After a participant agreed to participate in the study, he was given a web address where the instructions for the study were given. These instructions can be seen in section D.1.

After the participant had read the instructions and provided her username and password, she was directed to the questionnaire proper. In order to avoid duplication of answers, the questionnaire web pages were protected by a username and a password, which was given to each participant after they had agreed to participate in the study. The username and the password were only valid once, so that participants could not perform the study more than once.

Each participant was presented with 6 instructional settings, corresponding to one of the possible sequences introduced in section 9.2.1. They were asked to treat the instructional settings as sequential in time. For each instructional setting they were asked to infer the likely motivational state of a hypothetical student, and they were asked to

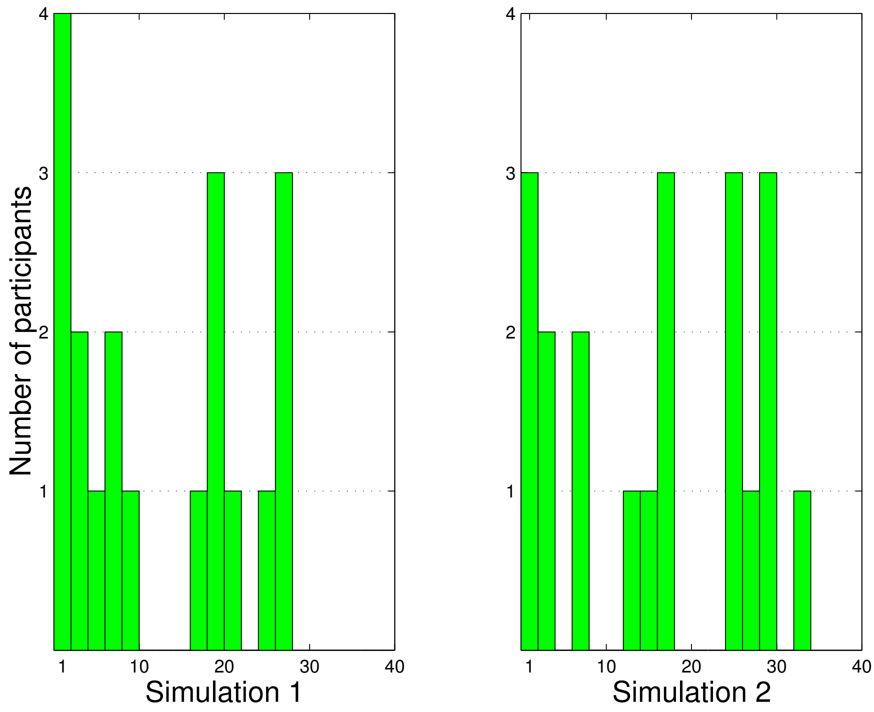


Figure 9.2: Teaching experience of participants in Evaluation Study.

comment on the importance they would give to each of the motivational factors if they had to take the next instructional step. If they did not have enough information to make a choice in any of the questions, they were asked to choose the *Don't know* option. They were also asked to give any extra comments and/or qualifications of their answers. Once they had finished answering all the questions, they were given the option of seeing the cumulative results.

9.5 Results

The raw data of the responses of the participants for both simulations can be seen in tables D.1 to D.4 in appendix D. Before we proceed to analyse in details the data obtained, we present in table 9.2 an example of the replies obtained in the motivational model question, and in table 9.3 an example of the replies obtained in the importance question.

SITUATION 1

	Very Low	Low	Average	High	Very High	Don't know
Satisfaction	0	0	1	4	14	0
Effort	0	1	2	8	7	1
Confidence	0	0	0	5	14	0
Cognitive Interest	0	0	1	6	8	4
Sensory Interest	0	0	3	5	5	6

Table 9.2: Example of results for motivation model question.

SITUATION 1

	Not Imp.	Sli. Imp.	Imp.	Very Imp.	Essential	Don't know
Satisfaction	0	0	6	11	3	0
Effort	0	0	5	12	3	0
Confidence	0	1	3	10	6	0
Cog. Interest	1	1	1	12	5	0
Sen. Interest	0	1	8	9	2	0

Table 9.3: Example of results for importance question.

9.5.1 Was it easy for participants to select a value?

For every question that the participants had to reply, they had a *Don't know* option that they could choose in the case that they did not have enough information to make a choice amongst the other values. In figure 9.3 we can see the distribution of *Don't know* replies for both the motivation model question and for the question regarding the importance of the motivational factors.

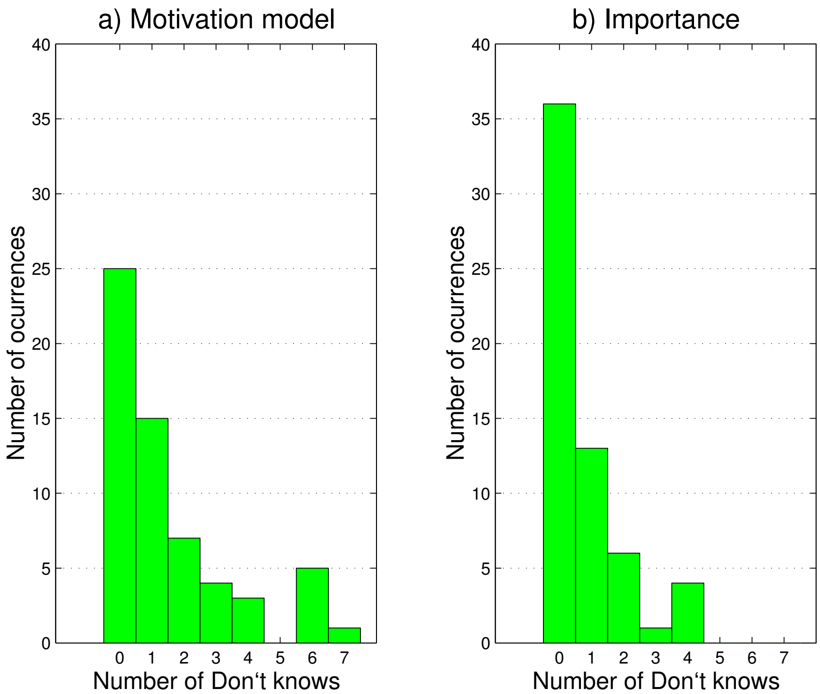


Figure 9.3: Distribution of *Don't know* replies for the two types of questions.

In figure 9.3a) we can see that no participant chose the *Don't know* option for twenty five out of the sixty individual questions[1] that participants replied regarding the motivational model of the hypothetical student. For one of the questions seven participants chose the *Don't know* option, and for five of the questions six participants chose the *Don't know* option (one of them being the question of Sensory Interest for the Situation 1, as seen in table 9.2).

[1]There were twelve situations, each of them with five motivational factors.

In what concerns the distribution of *Don't know* replies for the questions regarding the importance of the motivational factors, we can see in figure 9.3b) that for the majority of them (thirty six out of sixty) no participants chose the *Don't know* option.

From figure 9.3 it seems clear that most participants seemed to have enough information to reply to the large majority of the questions posed in our study, especially so for the questions regarding the importance of the motivational factors in their decision of the next instructional step to take.

Although the participants seemed to have more problems inferring the possible values for the motivational factors of the hypothetical student, as we can see in figure 9.4, this was due mainly to the *Sensory Interest* motivational factor.

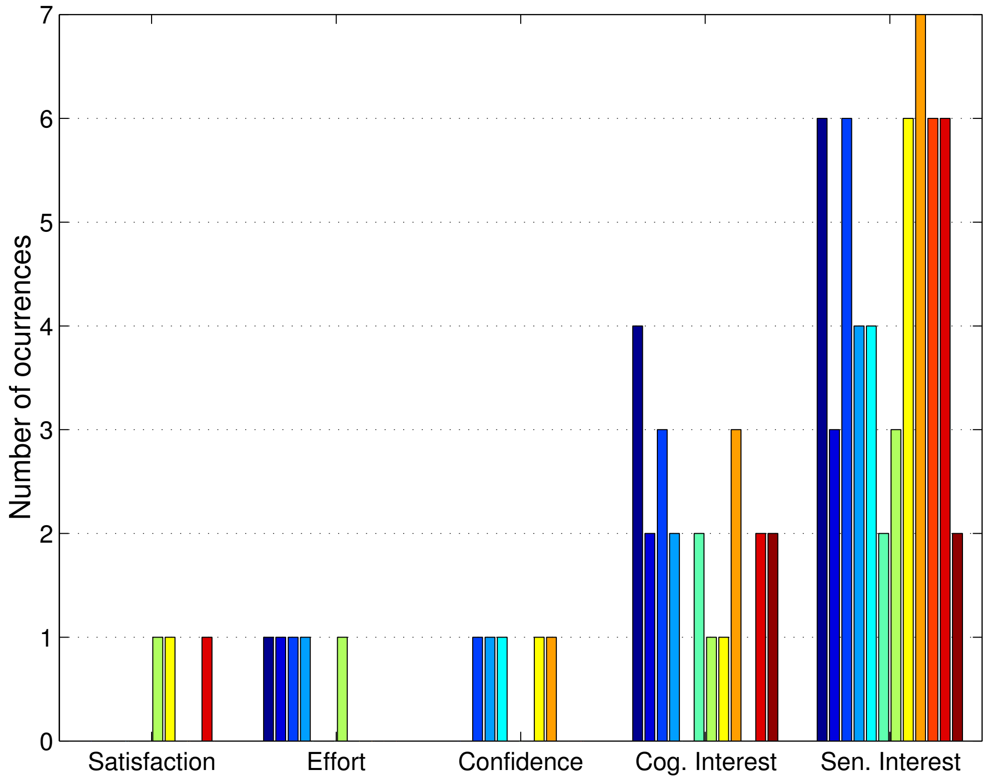


Figure 9.4: Distribution of *Don't know* replies for motivational model question by factors.

In figure 9.4 we show for each of the motivational factors the number of *Don't know* replies for each of the twelve instructional situations of our study. As we can see, the question regarding the *Satisfaction* of the student got almost no *Don't know* replies, except for three of the instructional situations, in which there was one *Don't know* reply.

On the other hand, we can see that for the question regarding *Sensory Interest* there were at least two participants who chose the *Don't know* option for all the instructional situations, and at least six participants chose *Don't know* for six of the instructional situations. This was somehow to be expected, as the nature of this evaluation is only textual, and there is very little information regarding sensorial input.

9.5.2 Did the participants agree on their inferences of the motivational model values?

Before we can compare the inferred values by MOODS with the inferred values by the participants of the evaluation, we need to analyse whether the participants agreed on their inferences. If due to any reasons (instructional situations too complex, motivational factors unclear, etc.) the participants did clearly disagree with each other on their inferences for any of the motivational factors, it would be meaningless to attempt to compare these with the inference made by MOODS.

If we consider the possible answers to each question to belong to a scale between 1 and 5[2], a measure of the agreement in the replies to a particular question is given by their standard deviation. Thus, the smallest the standard deviation, the bigger the agreement amongst all participants.

As an example of this, we see in table 9.4 the standard deviation values for the results of five fictitious questions with twenty participants. If ten of the participants were to choose the option *High* and the remaining ten were to choose the option *Very High*, the standard deviation would be 0.51. On the other side of the spectrum, if an equal number of participants were to choose each of the five possible options, the standard deviation would be high, at 1.45.

The judgement of what constitutes acceptable agreement amongst participants has to be a subjective one, but by looking at the examples in table 9.4 we can see that an

[2] 1 corresponding to the value *Very Low* and 5 to *Very High*.

σ	Very low	Low	Average	High	Very High
0.51	0	0	0	10	10
0.83	0	0	6	7	7
1.01	0	2	6	6	6
1.15	0	5	5	5	5
1.45	4	4	4	4	4

Table 9.4: Example of standard deviations as a measure of agreement.

standard deviation threshold of 1 seems to represent a set of question replies that show a clear grouping around one of the values in the scale.

If we look at the distribution of the standard deviation values for all the questions in our study, we can then have an understanding of the overall level of agreement amongst participants. We can see this for both simulations in figure 9.5.

We can see that the level of agreement for both simulations seems to be quite high. For example, in figure 9.5a) we see that for six of the questions the standard deviation was in the group (0.45, 0.55]. One of these is the question regarding *Confidence* in table 9.2, where 14 participants selected the option *Very High* and 5 participants selected *High*.

Overall twenty four out of the thirty questions for Simulation 1 and twenty eight of the thirty questions for Simulation 2 have a standard deviation of 1 or less, which shows that participants tended to agree on the general tendency of most of the motivational model questions.

9.5.3 How does MOODS compare with the inferences made by the participants?

As we have seen in the previous section, the participants of this study tended to mostly agree on the tendency of the values of the motivational factors. The mean of the values inferred by the participants for each of the motivational factors in both Simulations is given in figure 9.6, and the standard deviations in table 9.5.

As we can see in figure 9.6 a), the pattern of replies for each factor in Simulation 1

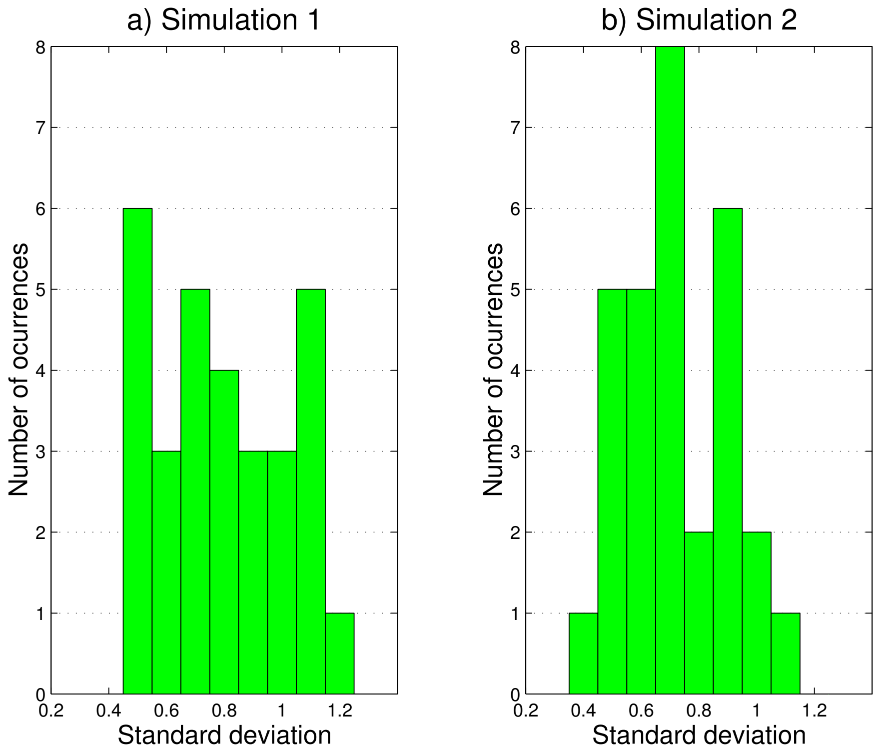


Figure 9.5: Distribution of standard deviations for questions in Simulations 1 and 2.

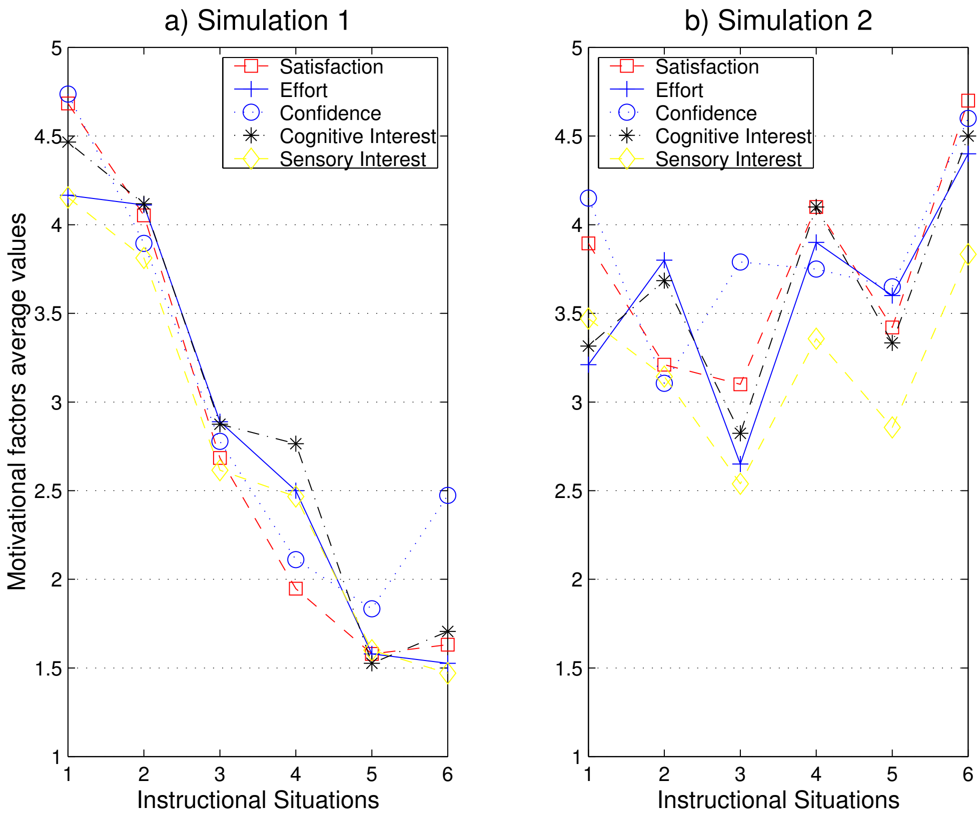


Figure 9.6: Average values of motivational factors for all instructional situations

is quite clearly downwards, which is in agreement with the intended characteristics set up for this simulation, as seen in section 9.2.1. The pattern of replies for Simulation 2 is, as it can be seen in figure 9.6 b), not so clearly defined, as the description of this simulation was not as simple as Simulation 1. We can also see that the range of values selected for the motivational factors in Simulation 2 is much smaller than in Simulation 1.

Now we can compare these values with the inferences made by MOODS for both Simulations, which are given in figure 9.7. For Simulation 1, we can see in figure 9.7 a) that the tendency of the values inferred by MOODS for each of the factors follows a similar pattern to those inferred by the participants of the evaluation. MOODS infers a high value for all the motivational factors in the first instructional situation, and there is a tendency towards lowering values during the Simulation, ending with low values

	Simulation 1						Simulation 2					
	Instructional Situations						Instructional Situations					
	1	2	3	4	5	6	1	2	3	4	5	6
Satisfaction	.58	.97	.67	.71	.51	.68	.74	1.03	.72	.64	.69	.47
Effort	.86	1.13	.76	.99	.51	.51	.63	.52	.93	.45	.68	.60
Confidence	.45	.88	.65	.68	.99	1.12	.67	.88	.86	.72	.59	.50
Cog. Interest	.64	1.05	.81	1.15	.51	.69	.89	.82	1.13	.55	.84	.71
Sen. Interest	.80	1.17	.87	1.06	.51	.80	.87	.53	.88	.50	.66	.99

Table 9.5: Standard deviations of motivational factors for all instructional situations

for nearly all the motivational factors.

For Simulation 2, and somehow similar to the inferences made by the participants, MOODS shows a much smaller variation of inferences, and the range of values given to each of the motivational factors is very limited. We can see that the only motivational factor whose value is changed during the Simulation is the one corresponding to *Satisfaction*. The Simulation 2 was obviously more difficult for MOODS, and there does not seem to be enough information to make many changes to the values of the motivational factors.

Difference of MOODS' and participants' inferences

In order to see how close the inferences made by MOODS were to those made by the participants we can analyse for each question the difference of the MOODS inference and the inferences made by the participants. For example, the question regarding *Satisfaction* for the instructional situation seen in table 9.2 has a mean value of 4.68. The value inferred for this factor in that instructional situation by MOODS is 4. Thus, the difference of the value inferred by MOODS and the mean of the value inferred by participants is -0.68.

If we calculate this difference for all the questions in both simulations we can draw a histogram showing the distribution of the differences of MOODS' inferences and the mean values of the participants' inferences. This can be seen in figure 9.8.

As we can see in figure 9.8 a), MOODS' inferences are most of the time very close

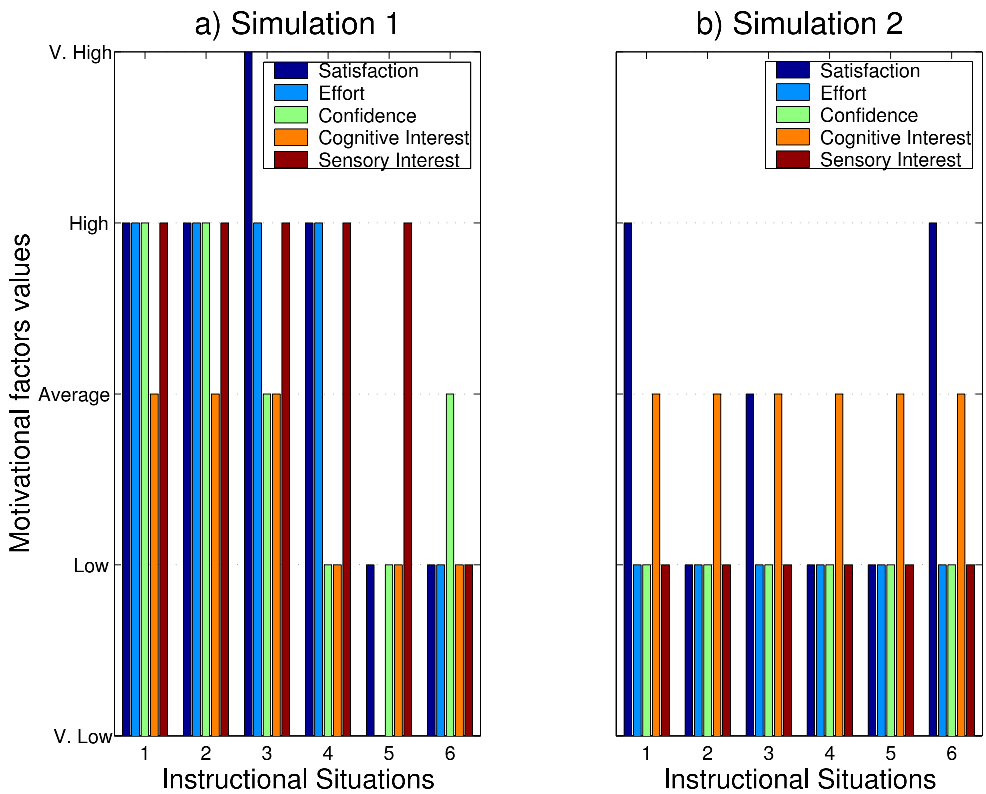


Figure 9.7: Values of motivational factors inferred by MOODS for all instructional situations

to the mean value of the participants' inferences in Simulation 1, the difference being less than one in twenty one of the thirty questions. On the other hand, the performance of MOODS in Simulation 2 was rather poor, and only in ten out of the thirty questions the difference of MOODS's inference was less than one.

A difference of less than one means that the inference of MOODS was the value for the motivational factor immediately below or above the mean of the values given by participants. For example, in the question regarding *Satisfaction* for the instructional situation seen in table 9.2, the mean value of the inferences made by the participants was 4.68, as seen above. A difference less than one from this value for MOODS means an inference of *High* (coded as 4) or *Very High* (coded as 5)[3].

[3]In the unlikely event that the mean was an integer number, a difference of less than one would imply an inference by MOODS of exactly the same value, but this was never the case in the data collected.

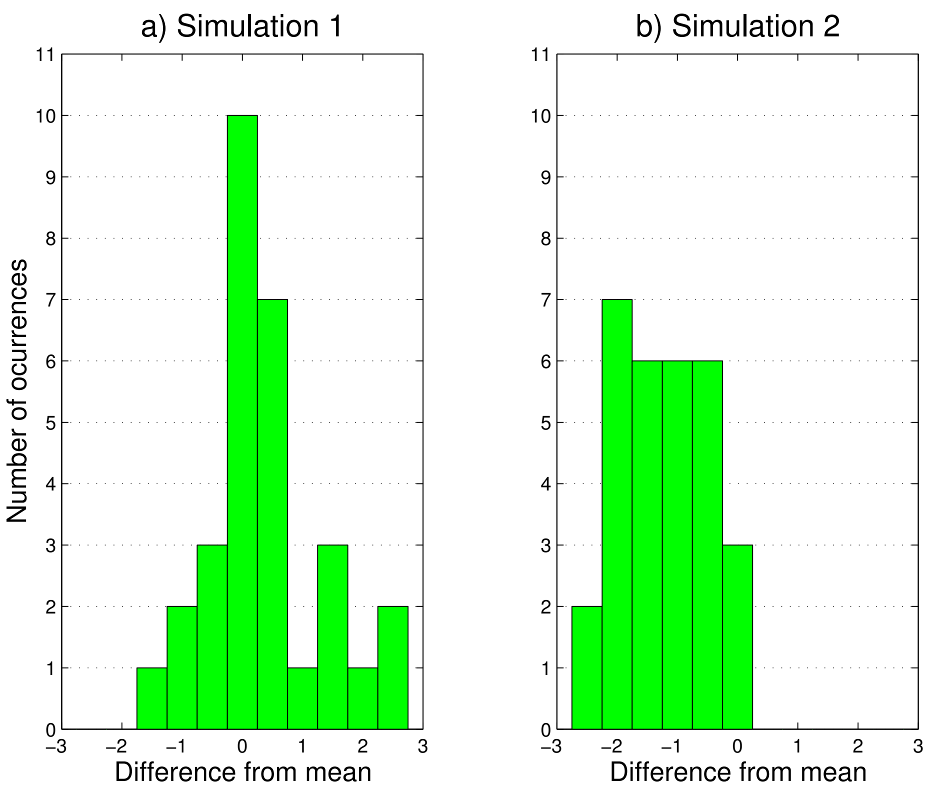


Figure 9.8: Distribution of the differences between the values inferred by MOODS and the average values by the study participants.

The probability of obtaining these results is analysed in table 9.6. If MOODS made the inferences about the values of the motivational factors randomly, we would expect a difference from the participant's mean of less than one in two out of five cases. Being then the null hypothesis that MOODS' performance is random, we can perform a chi-square *goodness of fit* test, the results of which are presented in table 9.6.

As we can see in table 9.6, the results for the Simulation 1 are statistically significant ($p<0.01$), but not for Simulation 2. As we saw in figure 9.7 b), MOODS was not able to make many changes to the values of the motivational factors across the instructional situations. At the same time, as we saw in figure 9.6, the participants' mean values for the motivational factors did not follow such a clear pattern as the values for Simulation 1. These two conditions might explain partially the bad results that MOODS shows for this Simulation.

	Simulation 1		Simulation 2	
	<1	≥ 1	<1	≥ 1
Obtained	21	9	10	20
Expected	12	8	12	18
		$p < 0.01$		$p = 0.46$

Table 9.6: Chi-square test

Difference in standard deviations of MOODS' and participants' inferences

In the previous section we have analysed the actual differences between the inferences made by MOODS and the mean of the participants' inferences for each question. It is also interesting to analyse the distance between MOODS' and the participants' inferences measured in terms of the number of standard deviations from the mean of the participants' replies.

By doing this, we can have a better picture of how close MOODS resembles a typical participant reply. For example, as we have seen in the previous section, the question regarding *Satisfaction* for the instructional situation seen in table 9.2 has a mean value of 4.68, and the value inferred by MOODS for this factor is 4. As we saw earlier, the difference of the value inferred by MOODS and the mean of the value inferred by the participants is -0.68. But the participants' replies for this question had a very small standard deviation, namely 0.58. This indicates that participants strongly agreed on the value for this factor. On the other hand, the replies for the factor *Effort* in the same situation had a larger standard deviation, namely 0.86. Therefore, if MOODS is to be compared with the performance of actual teachers, we should take this discrepancy into account.

We can do this by measuring the distance of MOODS' and the participants' replies in terms of the number of standard deviations away from the mean. Thus, the value inferred by MOODS for the *Satisfaction* factor is -1.17 standard deviations away from the mean of the participants' replies. If we calculate this distance for all the questions in both simulations we can draw a histogram showing the distribution of the differences in terms of standard deviations. This can be seen in figure 9.9.

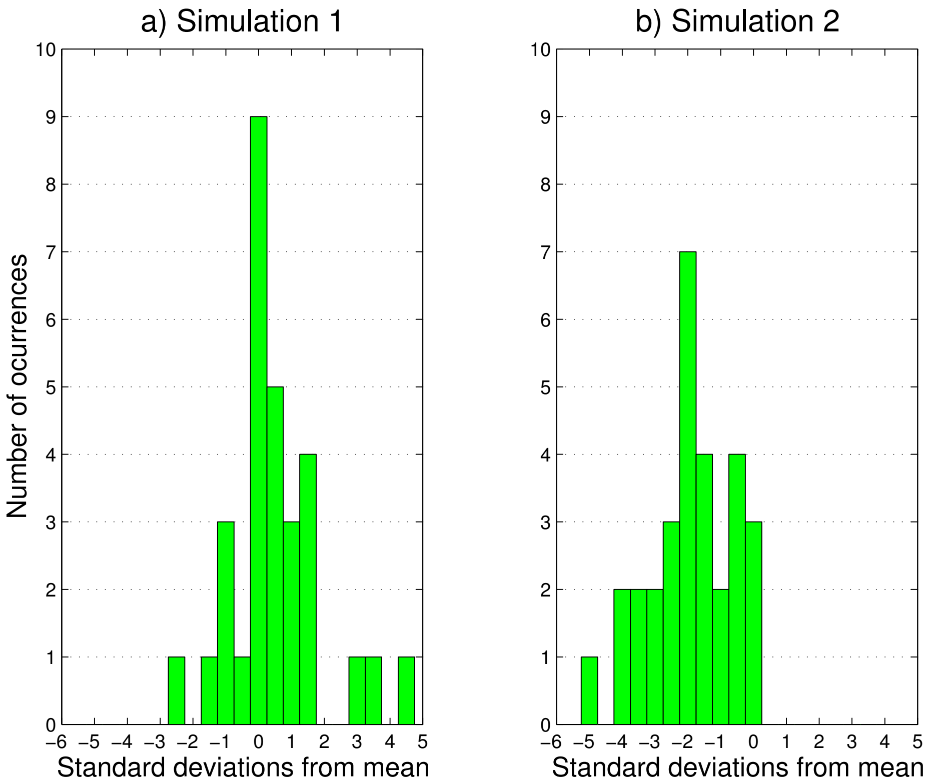


Figure 9.9: Distribution of the differences between the values inferred by MOODS and the average values of the study participants (in standard deviations from the mean).

Since the majority of the standard deviations of the replies of the participants are below one, the distances measured in terms of standard deviations (in figure 9.9) tend to be larger than the actual differences (in figure 9.8). Nevertheless, we can see that a similar pattern as that described in the previous section emerges. Figure 9.9 a) shows that MOODS' inferences were within one standard deviation from the participants' mean value in a large number of cases, namely eighteen out of thirty. In the case of the simulation 2, we see that this measurement also shows a poor MOODS performance, only eight out of the thirty values being within one standard deviation from the mean of the participants' replies.

9.5.4 How important were the motivational factors considered?

The importance that teachers would give to the values of certain motivational factors in order to plan the next instructional step to take can be of interest for the development of affective tutoring systems, mainly due to two reasons:

1. We should place special emphasis in diagnosing accurately those factors which are considered more important by teachers in their instructional planning. There is no point in diagnosing very carefully a factor that it is not going to play an important role in instructional planning.

2. We could use this information to improve the motivational planning component of a tutoring system by taking into account the relative importance of each factor.

As per the motivational model values, we analyse first whether participants tended to agree on the importance that they would place in each motivational factor. We show this by representing the distribution of the standard deviations of participants' replies in figure 9.10.

We can see that the agreement about the importance of the different motivational factors seemed to be greater for Simulation 2 than for Simulation 1. In Simulation 2 twenty nine of the thirty questions had a standard deviation of less than one, and in Simulation 1 twenty out of the thirty questions had a standard deviation of less than one.

In any case, it does not seem that we can extract too much information from the replies to the *Importance* question. If we look at figure 9.11, we can see the mean values of each of the motivational factors for all the instructional situations. For example, the first set of bars corresponds to the mean value of the *Satisfaction* motivational factor for the twelve instructional situations of the study (from Simulations 1 and 2).

Although there is some variation in the mean values in figure 9.11, we can see that most values tend to be around the *Very Important* mark for all factors and all instructional situations. In a sense this is not very informative, and it could not help us much to refine either the motivation diagnosis or the motivational planning components of an affective tutor.

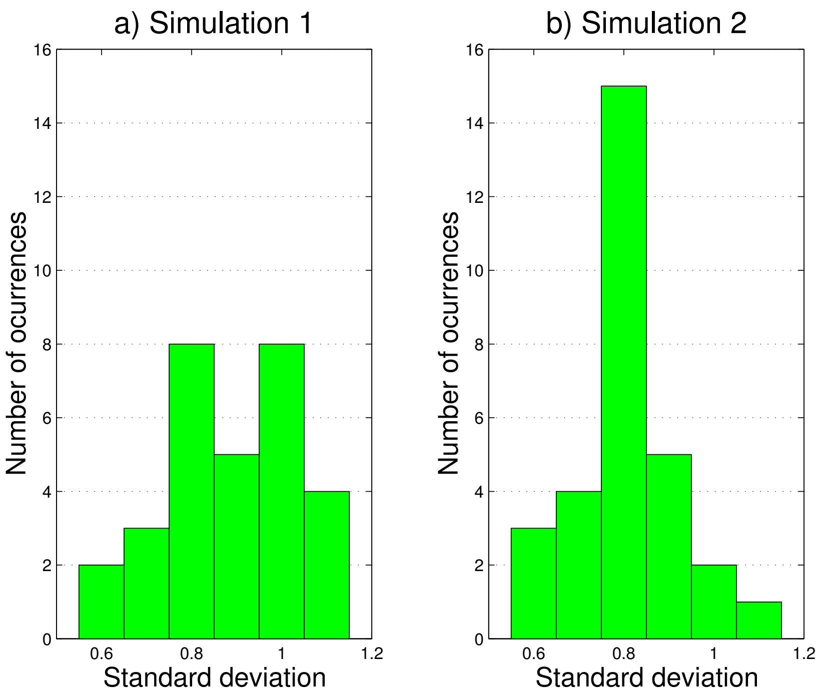


Figure 9.10: Distribution of standard deviations for *Importance* question in both simulations.

It is clear from participants' comments that the importance of these factors in order to plan the next instructional step was not always easy to judge. In many cases the difficulty in establishing the importance of a factor seems to be based on the lack of knowledge about the actual details of the instructional lesson or about the value of some of the motivational factors. For example, one of the participants commented:

> *I wish I knew more about his cognitive and/or sensory interest. I also wish I knew something about the demands he was faced with as effort is directly related to difficulty of which I have no indication. Just because he said he put in effort, I have no way to tell if he should procede to the next level of difficulty or if he should jump to a higher level. I also wish I knew more about the items he skipped as they would give me a better clue as to what should come next.*

This type of comment offers some indication about what other information would be necessary for participants to better judge the importance of each motivational factor

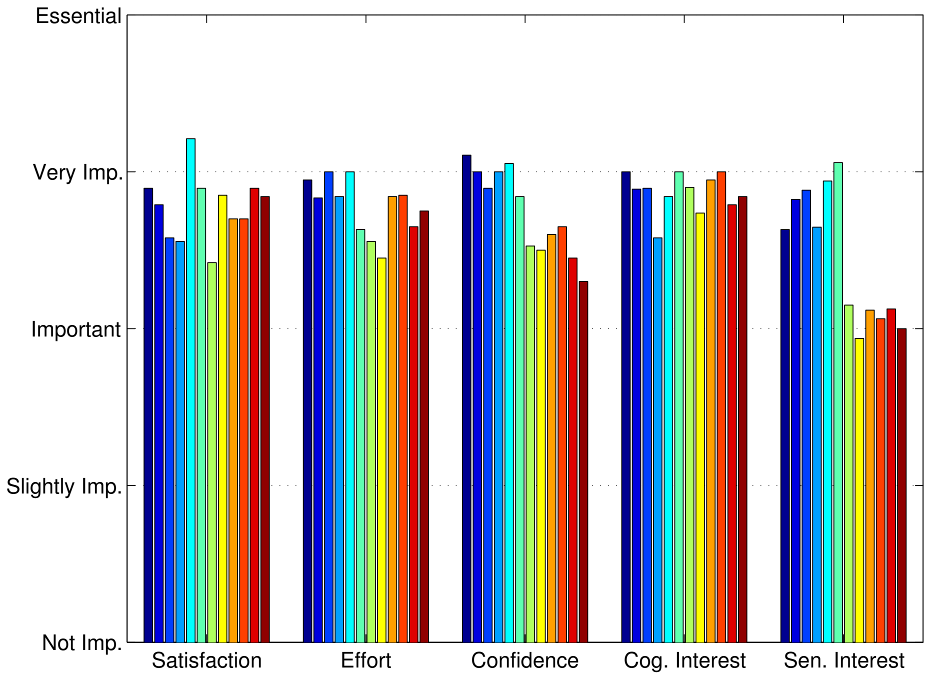


Figure 9.11: Average values of *Importance* question for all motivational factors and instructional situations.

for the planning of the next instructional step. But only seven comments about the importance of the motivational factors were recorded, so the scope for further analysis was very limited. Nevertheless, it is encouraging to see that the participants considered these factors very important in order to take the next instructional step, which seems to indicate that our model makes use of appropriate motivational constructs.

We also analysed whether the level of importance attached to the motivational factors had any effect on:

1. The value given to those factors by participants.

2. The level of agreement between participants.

3. The accuracy of the MOODS inference.

But, perhaps due partly to the very small variation in the values given to the *Importance* question, we found no correlation for any of these issues.

9.6 Discussion

In this chapter we have presented an evaluation of the motivation diagnosis knowledge implemented in MOODS. Evaluating motivation diagnosis knowledge is a very difficult task, as motivation is in itself a psychological construct for which a precise measurement tool can not be devised. Therefore, we evaluated MOODS by comparing the motivational values inferred by MOODS with the values inferred by human teachers under the same instructional situations.

As seen in chapters 7 and 8, MOODS was developed as a prototype affective tutor simulation, and as such, the evaluation presented in this chapter is based on the inferences made about the motivational state of a hypothetical student. That is, the participants in this evaluation were asked to imagine a situation in which a hypothetical student is using a computerised instructional system. They were presented with descriptions of a number of instructional settings and they were asked to infer the values of the motivational state factors. For each of these instructional settings we gathered the inferences about the motivational factors made by MOODS, and then we compared them with the inferences made by the human teachers.

By doing this type of evaluation, we managed to gather relatively quickly a large number of data. We prepared two different simulations, each of them consisting of six instructional settings. For each of these instructional settings we asked the participants to make inferences about the values of all the motivational factors represented in MOODS motivational model, and we also asked them to rate the importance that they would give to each of these factors if they had to decide which lesson to present next to the student.

We managed to get thirty-nine participants, nineteen of which followed the first simulation, and the remaining twenty followed the second simulation. By doing this evaluation as a web-based study, we managed to get a lot of data and we got the collaboration of very experienced teachers.

As a result of this evaluation, we learned a number of things about the feasibility of the motivation diagnosis approach implemented in MOODS:

- The participants of this study did largely agree on the value of the motivational factors for each of the instructional settings, which seems to indicate that the variables used to describe the instructional settings and the motivational state of the student are sufficient for the purpose of detecting the motivational state of the student.

- The performance of MOODS when compared to the inferences made by the human teachers was reasonably good for the first simulation, but not satisfactory for the second. We have seen in section 9.5 some of the possible reasons why the results for the second simulation were poor, but it was encouraging to see quite positive results for the first simulation. The motivation diagnosis knowledge in MOODS is far from complete, but we see that for some situations it seems to do a reasonable job at detecting the motivational state of the student.

- The data gathered regarding the importance that human teachers would place into each of the motivational factors modelled by MOODS was not very useful in terms of helping us refine the motivation diagnosis rules, but it showed that the model itself seems to be quite appropriate, in the sense that the participants consistently considered each of the factors quite important in order to decide which step to take next. Although a more careful study would be needed to see whether other factors would be considered equally important, none of the participants commented about this during the evaluation, which seems to indicate that our model offers enough information in order to plan the instruction taking into consideration motivational issues.

Overall, we see that this evaluation allowed us to gather relatively quickly a large amount of data about inferences made by human teachers which we then compared with the inferences made by MOODS. The results, though not extremely good, were encouraging. Nevertheless, we have to remember that these data were based on a simulation of an instructional interaction, and not on a real interaction. Due to lack of time and resources, it was not possible as part of this research to perform an evaluation of the motivation diagnosis knowledge implemented in MOODS based on real instructional interactions of students with an ITS. But it would be interesting to perform an alternative evaluation, in which an actual ITS is augmented with the motivational components implemented in MOODS. This and other possible further work is presented in greater detail in chapter 10.

Chapter 10

Conclusions and Further Work

10.1 Summary of research and findings

As we saw in chapter 2, there is recently a high interest in the ITS community in research that takes into consideration affective issues. But this new and exciting research area covers a vast number of topics, and therefore all the related research deals by necessity with a very specific and small issue.

In this sense, the research presented in this dissertation is no different from previous research, as we have focused on a very concrete and specific issue, i.e. the detection of a student's motivational state in terms of a basic motivational model, based on self-report and the student's interaction with an ITS.

The detection of a student's motivational state has been virtually unexplored in ITS research, except for the work of del Soldato (1994). But, as mentioned in chapter 2, her work focused mainly on motivational planning and therefore the motivation detection components of her work were somehow speculative and not thoroughly evaluated.

In our research, we focused specifically on the problem of detecting the motivational state of a student interacting with an ITS . At the same time we argued that the available theories of motivation in education are not specific enough in order to rationalise and implement knowledge about motivation detection in an ITS . Therefore we conducted a number of empirical studies to inform the design of a prototype ITS that detects the motivational state of a student.

Given the confusion that sometimes arises in relation to affective terminology, we started in chapter 2 by briefly explaining the different meanings of the terms ‘affect’, ‘emotions’ and ‘motivations’. We then reviewed the research related to motivation in education and specifically to motivation in ITSs . Although our work focuses on the psychological construct of motivation, we also reviewed some of the work related to emotions, since both constructs are obviously related and they influence each other.

We saw that there have been a variety of different approaches to issues like the detection by a computer of a person’s emotions, such as: interpreting his physiological data (blood volume, respiration, etc.); recognising his facial expressions; etc. But we also saw that research dealing explicitly with motivation in ITSs was very limited and that the issue of motivation detection was worth exploring in order to create tutoring systems that are able to better empathise with the student.

Thus, we presented in chapter 3 an outline of the design of a tutoring system that could detect the motivational state of a student. We presented a model of motivation based on a number of theories of motivation in education which we would use to represent the motivational state of a student, and introduced the methods for motivation detection that we would explore in our dissertation.

In chapter 4 we focused on self-report as a method to diagnose a student’s motivational state and we discussed an empirical study that we performed in order to:

1. investigate whether self-report could really be a viable choice as a motivation diagnosis technique;
2. and if this was the case, to find out which approach to self-report should be the most appropriate to present to a student.

We found that self-report was in general well received by the students, but analysis of the data recorded suggested a number of modifications to the original self-report interface in order to make it more suitable for the detection of a student’s motivational state.

We also found that to rely exclusively on self-report would not be appropriate, as there are a number of situations when the student might not update the self-report facilities as frequently as we would require in order to adapt the instruction appropriately to his motivational state.

As a result of this, we set out to study the possibility of formalising knowledge about motivation diagnosis, which could then be implemented in a tutoring system. As mentioned above, previous work in this area seemed to be based mainly on intuition, and the relevant theories of motivation in education seemed to offer suggestions which were vague, too general, or too complex to implement in a tutoring system.

Therefore, we performed another empirical study in which we presented participants with a previously recorded instructional interaction, and we asked them to try to infer the motivational state of the student at certain times during the interaction. We also asked them to formalise their reasoning in order to create a set of motivation diagnosis rules that could be easily implemented in a tutoring system. This study was reported in chapter 5.

As a result of this study, and contrary to the participants' expectations, we collected a large number of rules regarding the detection of a student's motivational state. Nevertheless, these rules had to be validated, as they came from a small population of participants and a number of errors could have been introduced in the rules, for example when transcribing the comments of the participants. Thus, we designed another empirical study to validate these rules, which we described in chapter 6.

In this study we presented participants with an instructional interaction context and asked them to rate the rules that could be applied under those conditions. By doing this, we found which rules from those elicited in the previous study were generally accepted as valid by teachers, and which ones were not. We found that a very large number of the rules (forty-one out of the total of sixty-one) were accepted as valid[1]. We also gathered a large number of comments from the participants, which could help us to better understand the motivational detection knowledge of the teachers and to further improve the set of detection rules.

In order to test this approach to motivation diagnosis in an ITS we developed MOODS, a simulation of a tutoring system, in which we implemented the knowledge elicited during the previous studies, and also the motivational planning rules developed

[1]I.e. participants showed a statistically significant preference ($p<0.01$) for the *accept* category.

by del Soldato (1994). MOODS was developed as a simulation of a tutoring system and not as an actual ITS for three main reasons:

1. Simplicity of development
2. Simplicity of evaluation
3. Generality

The overall structure of MOODS was presented in chapter 7 and a sample interaction with it was given in chapter 8. As we saw in chapter 8, MOODS is a simple system which represents an embodiment of the motivational knowledge described above. It is not an actual tutoring system, but it allows us to quickly test and evaluate the motivation diagnosis knowledge implemented in it by simulating an instructional situation and querying the system about the likely changes in the motivational state of a hypothetical student.

In order to evaluate the motivation diagnosis knowledge implemented in MOODS, we performed another study, which is described in chapter 9. Given the inherent difficulties of this type of evaluation, we evaluated the motivation diagnosis knowledge by developing two different instructional simulations, each consisting of six simulated instructional units, and for each instructional unit we asked the study participants to infer the motivational state of a hypothetical student. Then we compared the values inferred by the human teachers to those inferred by MOODS.

The results of this evaluation were not as good as we had hoped, but they were reasonably encouraging. For one of the simulations the results produced by MOODS were very similar to the ones inferred by the teachers. On the other hand, the values inferred by MOODS in the second simulation were not really similar to the values inferred by the participants.

Nevertheless, the results of the evaluation were encouraging as they seemed to indicate amongst other things that:

1. The factors used in MOODS to describe an instructional setting convey enough information to make valid inferences about the motivational state of the student.

2. The model used in MOODS to represent the motivational state of the student is considered by teachers important information in deciding the next instructional step to take.

3. The motivation diagnosis knowledge of MOODS seems to be valid in at least some instructional interactions.

Behind the inference mechanism built in MOODS lays the motivation model introduced in chapter 3. As mentioned in chapter 3, many other factors from other accounts of motivation could have been included in our motivation model, but we suggested that our model represented a useful set of the main important characteristics of a student's motivation, while at the same time being small enough so that it could be incorporated easily into a tutoring system.

We believe that this model has stood up well to the different empirical studies presented in this dissertation:

- Except the "Relevance" factor which was eliminated from the model after performing the self-report study (see chapter 4), the distinction between traits and states and the actual meaning of each of the factors proved easily understandable by all the participants of the mentioned empirical studies.

- At the same time, these factors seem to be quite appropriate when talking about a student's motivational state. This is the case, since the participants of the three studies were able to relate their interaction and the student's motivational state to these factors.

- As shown in the evaluation of MOODS in chapter 9, having information about the values of these factors was considered important by the participants in order to plan the next instructional step.

Nevertheless, this motivation model should not be considered as the only or the ultimate way to represent a student's motivational traits and state. The model seemed to prove appropriate for a generic description of a student's motivation under a generic instructional experience. But if the model were to be applied to an actual tutoring system, it would be necessary to study how it should be adapted to the given domain and students. That is, we believe that the model could serve as an starting point, but refinement would be needed to adapt it to a particular situation. For example, it is likely that the "Fantasy" factor should be further refined when dealing with children in order to accommodate their fantasy needs.

All this considered, we believe that the approach used in MOODS represents an encouraging step in the direction of creating tutoring systems that can monitor the motivational state of the student, but many improvements could be made to it in order to improve its efficiency. In the following sections we look at some of the problems with the approach followed in the system evaluation and also at some interesting further work that could help us to improve the motivation diagnosis capabilities of a tutoring system.

10.2 Some criticisms

There is no doubt that, if given a second chance, some aspects of this thesis should perhaps be dealt with differently. The issues researched in this dissertation are very complex, and some of the methodological decisions could be criticised. In this section we comment on some of these issues.

Is the motivation model appropriate? A crucial point throughout the dissertation is the use of the motivation model introduced in chapter 3. The design of this model was based on theories of motivation in Education, and as such, we believe that it represents a useful model of a student's motivation. Nevertheless, it is true that this model is generic in nature. We attempted to capture in it the common motivational factors that could be applied to every domain and every student. Therefore, we could wonder if the model is so generic that it would be unusable in a real tutoring system.

As pointed out in section 10.1, we believe that the motivation model represents a useful approach to motivational modelling, but it would have to be adapted to a given domain and target students if it were to be applied to an actual tutoring system.

Should we have concentrated on a particular domain and target students? As mentioned above, the motivation model (and all the dissertation) assumes a generic instructional domain and a generic student. Thus, we studied issues of motivation detection for all possible types of tutoring systems. But as per the motivation model criticism above, one could wonder if it would have been more informative to focus into a particular domain and target students.

One possibility would have been to use an existing tutoring system, and make use of its instructional domain and target students to develop the motivation model and the motivation detection rules. Probably this approach would have provided more detailed information about motivation detection in that domain and for those students, but generality would have suffered. We believe that it was appropriate to perform first a more generic investigation, and to consider a more specific one as possible further work.

Was the methodology used for the self-report study adequate? As explained in section 4.3, 18 students volunteered to participate in the self-report study, and we divided these students into six groups corresponding to the six developed instructional paths. We considered that it was important to have participants in each of the instructional paths, but looking at the collected data retrospectively, we see that the variations from one path to another are not so great. Thus, we could perhaps argue that it would have been more informative to develop fewer instructional paths, so that more participants could follow each of them. This design would have probably provided similar results in terms of the qualities of self-report, and at the same time it would have made possible to perform statistical analysis in the data for each of the instructional paths.

Was the knowledge elicitation approach appropriate? As mentioned in section 5.1, the approach we followed in order to elicit the knowledge about motivation detection and the development of MOODS can be compared to that used in a knowledge engineering process. As we saw in section 5.1 we can relate the approach followed in this dissertation to the five common steps in knowledge engineering:

1. We performed Knowledge Acquisition in the first part of the motivation diagno-

sis study by recording the comments made by the participants while watching the interaction of a student with an instructional system and by creating a first approximation of the rules representing the knowledge about motivation detection.

2. We performed Knowledge Representation in the second part of the motivation diagnosis study by developing a final set of motivation diagnosis rules based on the knowledge previously acquired.

3. We performed Knowledge Validation in the Motivation Diagnosis Validity Study presented in chapter 6.

4. We performed Inference by developing MOODS.

5. We performed a basic Explanation and Justification step by developing MOODS with a facility to give basic explanations of its reasoning.

As we can see, the steps are similar to those used in Knowledge Engineering, but the actual techniques we used could be improved on. By reviewing some of the related work (see, for example Boose, 1989; Cooke, 1994; Gaines, 1987; Turban, 1992), we can see that there are many techniques for knowledge acquisition, and it is not trivial which one to select in a particular situation, as this is a field that is still considered more an art than a science.

The techniques for knowledge acquisition can be broadly categorized as follows:

- **Manual methods**, in which a knowledge engineer establishes some kind of interview with the subject expert in order to elicit the expert knowledge.

- **Semiautomatic methods**, in which experts can build knowledge relying little or not at all on the knowledge engineer.

- **Automatic methods**, such as induction, in which a knowledge base can be built with little or no need for either a Knowledge Engineer or an expert. In this type of methods, an expert is normally needed only for validation of the knowledge.

Manual methods include:

- **Interviews**, which can be unstructured or structured.
  - Unstructured interviews are informal and normally used as a starting point, since they require little planning and offer a quick way of understanding the structure of the problem domain.
  - Structured interviews are more formal, and they are well organized, following a systematic approach.
- **Tracking methods**, which can be informal (Observations) or formal (Protocol Analysis).
  - Observations involve the knowledge engineer observing the expert work in the field, looking at the performance of the task.
  - Protocol Analysis is similar to interviewing but more formal. The expert is asked to carry out a task while thinking out loud about the process itself.

For the studies performed in this dissertation we have followed a mixture of *Protocol Analysis* and *Structured interviews*. These approaches are very common for knowledge elicitation, but knowledge elicitation using these techniques is very slow and the existence of a knowledge engineer can actually cause more problems than it solves (Gaines, 1987).

Thus, we should consider that perhaps we could have made use of a computer-based acquisition tool (see, for example Boose, 1989), which could perhaps facilitate the quick elicitation of more and deeper knowledge about motivation detection.

Was the approach used for the evaluation valid? The evaluation presented in chapter 9 was (partly due to pragmatic reasons) based on simulations of instructional interactions. As we have seen, this allowed us to perform the evaluation of MOODS quickly and with a large number of participants, but it could be argued that the simulations were too artificial and that it would be necessary to implement the same techniques into a real ITS in order to perform another evaluation. We believe that the evaluation performed was useful and helped to show that the techniques developed in MOODS were promising, but it would be worthwhile to perform a larger study, in which the techniques implemented in MOODS are incorporated into a real ITS and evaluated again for comparison.

At the same time, the instructional settings used in the system evaluation are represented only by a few characteristics. The need to present the instructional interactions in a concise way to the participants meant that not all possible variables could be used. As we explained in section 9.2.1, we represented the instructional settings with those variables that were used more often in the motivation diagnosis rules. In this way we tried to minimise the effect of not including all the necessary variables in the description of the instructional settings. But if further evaluation of the system were to be carried out, it would be desirable to find a way of representing each instructional setting with all the instruction characteristics used in MOODS.

It is also interesting to note that the results for the question regarding the importance of the motivational factors (see section 9.5.4) did not offer a great deal of information, as the importance of the motivational factors was not always easy to judge. In retrospect, it is clear that it would be necessary to design another way to ask participants about the importance of the motivational factors, in order to obtain more meaningful results to this question.

10.3 Further work

The work presented in this dissertation does not represent the goal of motivation diagnosis research, but rather a stepping stone from which we could perform further research. In this section we present a small list of suggestions of further work that could be carried out based on this research.

1. **Further development of MOODS as a testing tool.** MOODS could be improved in order to develop a better tool to test, evaluate and improve the implemented motivation diagnosis knowledge. In its current version, MOODS' functionality is very limited, and it would be useful to add a number of facilities. For example, we could expand MOODS so that individual rules can be edited, ignored, added, etc. We could also add facilities for deciding which performance variables are taken into account. For instance, we could decide that we are not in-

terested in considering the variable "Hesitation" when firing the diagnosis rules, etc.

2. **Study the possibility of other modalities of self-report.** The self-report study we performed made use of sliders as a representation of the motivational state of the student. It would be interesting to look at other modalities of representation of the motivational variables, for example graphics, 'emoticons', etc. Would different modalities influence the acceptance towards using the self-report facility? Would the personality traits (for instance, fantasy) influence the preference towards one type of motivational model representation?

3. **Evaluation of MOODS compared with student's self-report.** The evaluation we performed compared the inferences made by a group of teachers with the inferences made by MOODS. Despite some of the drawbacks of the self-report method mentioned in chapter 4, it would be very interesting to perform a further study in order to compare the inferences made by MOODS with the self-report made by students. Do these differ greatly? Does MOODS predict better students' self-report or teachers' inferences?

4. **Comparison of self-report and teachers' inferences.** Another interesting analysis would be to compare the self-report inferences with those made by teachers. How do they differ? Is it possible to predict some of the differences between teachers' inferences and the self-report, given the personality traits of students?

5. **How to rate the importance of different diagnosis channels?** As we explained in chapter 8, the combination of the self-report readings and the inferred values was implemented ad hoc in MOODS. If we consider that an ITS could use as well other channels of motivation/emotion detection (such as recognition of facial expressions, voice intonation, etc.), it becomes increasingly important that we investigate how we could combine the information from all these channels. Which channel should we trust in case of discrepant readings? Are some channels redundant and provide similar readings?

6. **Study the possibility of integration of a motivational model and a model of**

emotions. As reviewed in chapter 2, there has been some research on developing ITSs that can detect the emotional state of the student. The emotions that are normally mentioned in this type of research refer to frustration, anxiety, etc. Given the existing relation between emotions and motivation, it would be useful to study how these models could be amalgamated in order to create a more accurate representation of the student. How do emotions influence motivation? Would it be useful for an ITS to model basic emotions of the student (such as happiness, sadness, etc.), or do these occur too seldom during an instructional situation to be useful?

7. **Generalisation of motivation diagnosis rules.** As a result of the empirical studies presented in this dissertation, we have a variety of rules for motivation detection, but these do not form a comprehensive and/or cohesive representation of motivation diagnosis knowledge. As explained in chapter 8, in the prototype developed we considered the applicable rule to be the one for which the highest number of preconditions hold in the current instructional situation. This was implemented somehow ad hoc in order to allow for some generalisation of the rules. An important development of this work would be to attempt to create a more generalisable representation of the motivation diagnosis knowledge. A possibility could be to develop a Bayesian model of motivation detection based on the currently implemented knowledge. This type of model has been used in some of the research on modelling emotions and it would be interesting to see whether this type of representation could increase the efficiency of the motivation detection capabilities of MOODS.

10.4 Conclusions

The detection of the motivational state of a student is a crucial step if we want to create tutoring systems that are able to empathise with their students. In this dissertation we have focused on this issue, paying special attention to empirical information gathered via a number of studies.

By relating the goals of this thesis (as mentioned in section 1.3) with the results of the various studies and evaluation, we can see that the main contribution of this work has been threefold:

1. To explore issues of diagnosis of a student's motivation. We have reviewed a large body of research work on issues of detection of emotions and motivation in chapter 2. We have explored in detail the use of self-report and we have elicited and validated a large set of knowledge-based rules for motivation detection.

2. To explore ways of formalising knowledge about affective instruction. We have devised and carried out three different empirical studies (presented in chapters 5, 6 and 9) by which we elicited the knowledge-based rules mentioned above.

3. To explore the issue of affective tutors evaluation. As mentioned above, the evaluation of the motivation diagnosis knowledge acquired via the empirical studies was a difficult task. By preparing an evaluation of an instructional interaction simulation, we were able to evaluate MOODS' performance against the performance of a large number of teachers.

At the same time, we developed a prototype tutoring system simulation in which we implemented all the knowledge elicited via the empirical studies. The evaluation carried out in chapter 9 did not show perfect results for our motivation detection engine, and as we have pointed out in section 10.3, there is still a lot of work that could be done in order to improve the motivation detection knowledge in MOODS. However, in this dissertation we have presented what we believe is a promising step into creating tutoring systems that care!

Bibliography

Affect in Interactions (1999). International Workshop on Affect in Interactions. Towards a New Generation of Interfaces (Annual Conference of the EC I3 Programme). Available on the World Wide Web at: `http://gaiva.inesc.pt/i3ws/i3workshop.html` (October 1, 2001). Revised full versions of several papers and various selectively solicited papers by key people in the field can be found in (Paiva, 2000).

Allport, G. (1992). Towards a computational theory of dialogues. CSRP Report 92-3, University of Birmingham. School of Computer Science. Cognitive Science Research Centre.

Ames, R. E. and Ames, C., editors (1984). *Research on Motivation in Education. Volume 1. Student Motivation*. Academic Press, Inc., Orlando, Fla; London.

Arshad, F. (1990). *The Design of Knowledge-based Advisors for Learning*. Ph.D. thesis, School of Education, University of Leeds, UK.

Attitude, Personality and Emotions in User-Adapted Interaction (2001). 2nd Workshop on Attitude, Personality and Emotions in User-Adapted Interaction in conjunction with User Modeling 2001. Available on the World Wide Web at: `http://aos2.uniba.it:8080/ws-um01.html` (April 8, 2002).

Ball, S., editor (1977). *Motivation in Education*. Academic Press, New York.

Bares, W. H., Zettlemoyer, L. S., and Lester, J. C. (1998). Habitable 3D learning environments for situated learning. In Goettl *et al.* (1998), pages 76–85.

Bauer, M., Gmytrasiewicz, P., and Vassileva, J., editors (2001). *Proceedings of the Eigth International Conference on User Modeling in Sonthofen, Germany*, volume 2109 of *Lecture Notes in Computer Science*, Wien, New York. Springer-Verlag.

Beck, J., Stern, M., and Woolf, B. P. (1997). Cooperative student models. In du Boulay and Mizoguchi (1997), pages 127–134.

Binsted, K. and Luke, S. (1999). Character design for soccer commentary. In M. Asada and H. Kitano, editors, *RoboCup-98: Robot Soccer World Cup II*, volume 1604, pages 22–34, Heidelberg. Springer-Verlag. Available on the World Wide Web at: `http://www.cs.gmu.edu/~sean/papers/` (October 1, 2001).

Boose, J. (1989). A survey of knowledge acquisition techniques and tools. *Knowledge Acquisition*, **1**(1), 39–58.

du Boulay, B. and Luckin, R. (1999). It ain't what you learn but the way that you learn it. *Computers & Education*, **33**(2-3), 209–215.

du Boulay, B. and Mizoguchi, R., editors (1997). *Proceedings of the Eigth World Conference on Artificial Intelligence in Education*, Amsterdam. IOS Press.

Briggs, P., Burford, B., and Dracup, C. (1996). Self confidence as an issue for user-modeling. In *Proceedings of the Fifth International Conference on User Modeling in Kailua-Kona, Hawaii*, pages 179–186, Boston, MA. User Modeling, Inc.

Brna, P., Paiva, A., and Self, J., editors (1996). *Proceedings of the European Conference on Artificial Intelligence in Education*, Lisbon. Colibri.

Bull, S. (1997). *Collaborative Student Modelling in Foreign Language Learning*. Unpublished PhD Thesis, Department of Artificial Intelligence, University of Edinburgh.

Bull, S., Pain, H., and Brna, P. (1994). Student modelling beyond domain knowledge. In *Proceedings of the Fourth International Conference on User Modeling*, Hyannis, Massachusetts, USA. The MITRE Corporation.

Cahn, J. E. (1990). The generation of affect in synthesized speech. *Journal of the American Voice I/O Society*, **8**, 1–19.

Cerri, S. A., Gouardères, G., and Paraguaçu, F., editors (2002). *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems*, volume 2363 of *Lecture Notes in Computer Science*, Berlin. Heidelberg. Springer.

Clifford, M. M. (1990). Student motivation and academic achievement. In Vedder (1990), chapter 4, pages 59–89.

Colby, K. M. (1981). Modeling a paranoid mind. *The Behavioral and Brain Sciences*, **4**(4), 515–560.

Conati, C. and Zhou, X. (2002). Modeling students' emotions from cognitive appraisal in educational games. In Cerri *et al.* (2002), pages 944–954.

Cooke, N. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, **41**, 801–849.

Crowley, C. (1996). TkReplay. Version 1.1. Available on the World Wide Web at: `http://www.cs.unm.edu/~crowley` (November 18, 1998).

Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*. Behavioral Science Series. Jossey-Bass Publishers, San Francisco.

Damasio, A. R. (1996). *Descartes' Error. Emotion, Reason and the Human Brain*. Papermac, London. First Published as a Grosset/Putnam Book 1994 by G.P. Putnams Sons, New York.

Davis, F. (1976). *La Comunicación No Verbal*, volume 616 of *El Libro de Bolsillo*. Alianza Editorial, Madrid, Spain. Translated by Lita Mourglier from: "Inside Intuition - What we Knew About Non-Verbal Communication"; McGraw-Hill Book Company, New York.

Dyer, M. G. (1983). *In-depth Understanding: A Computer Model of Integrated Processing For Narrative Comprehension*. MIT Press, Cambridge, Mass.; London.

Ekman, P. and Friesen, W. (1977). *Manual for the Facial Action Coding System*. Consulting Psychologists Press, Palo Alto.

Ekman, P. and Rosenberg, E. L., editors (1997). *What the face reveals : basic and applied studies of spontaneous expression using the facial action coding system (FACS)*. Oxford University Press, New York, Oxford.

Elliott, C. D. (1992). The affective reasoner: a process model of emotions in a multi-agent system. Technical Report 32, Northwestern University. The Institute for the Learning Sciences.

Emotion-based Agent Architectures (1999). Proceedings of the Workshop on Emotion-Based Agent Architectures (EBAA'99). Third International Conference on Autonomous Agents (Agents'99), Seattle, WA. Available on the World Wide Web at: `http://www.ai.mit.edu/people/jvelas/ebaa.html` (October 18, 1999).

Emotion in HCI (1999). A British HCI Group one-day meeting in conjunction with University College London on "Affective Computing: The Role of Emotion in HCI". Available on the World Wide Web at: `http://www-users.york.ac.uk/~am1/affective.html` (April 8, 1999).

Emotional and Intelligent (1998). *Proceedings from the Symposium 'Emotional and Intelligent: The Tangled Knot of Cognition'. Technical Report FS-98-03*, Orlando, Florida. AAAI Press.

Essa, I. and Pentland, A. (1997). Coding, analysis, interpretation and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(7), 757–763.

Essa, I. A. (1995). *Analysis, Interpretation and Synthesis of Facial Expressions*. Ph.D. thesis, MIT Media Lab, Cambridge, MA.

Frasson, C., Gauthier, G., and Lesgold, A., editors (1996). *Proceedings of the Third International Conference on Intelligent Tutoring Systems*, Berlin. Springer-Verlag.

Gaines, B. (1987). An overview of knowledge acquisition and transfer. *International Journal of Man-Machine Studies*, **26**(4), 453–472.

Gardner, R. C. (1985). *Social Psychology and Second Language Learning: the Role of Attitudes and Motivation*. London: Edward Arnold.

Georgouli, K. (2002). The design of a 'motivating' intelligent assessment system. In Cerri *et al.* (2002), pages 811–820.

Goettl, B. P., Halff, H. M., Redfield, C. L., and Shute, V. J., editors (1998). *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, volume 1452 of *Lecture Notes in Computer Science*, Berlin. Heidelberg. Springer.

Goleman, D. (1996). *Emotional intelligence: Why it can Matter more than IQ*. Boomsbury, London. First published by Bantam Books, New York, 1995.

Greer, J. E. and McCalla, G. I., editors (1994). *Student Modelling: The Key to Individualized Knowledge-Based Instruction*, volume 125 of *NATO ASI series. Series F, Computer and Systems Sciences*. Springer Verlag in cooperation with NATO Scientific Affairs Division, Berlin; London.

Hietala, P. and Niemirepo, T. (1997). Collaboration with software agents: What if the learning companion agent makes errors? In du Boulay and Mizoguchi (1997), pages 159–166.

Hioe, W. and Campbell, D. J. (1988). An expert system for the diagnosis of motivation-related job performance problems: Initial efforts. DICS publication No. TRA3/88, Department of Information Systems and Computer Science, National University of Singapore.

Horvitz, E. and Paek, T. (1999). A computational architecture for conversation. In Kay (1999), pages 201–210. Available on the World Wide Web at: `http://www.cs.usask.ca/UM99/` (May 24, 2000).

Issroff, K. (1996). *Investigating Computer-supported Collaborative Learning from an Affective Perspective*. Unpublished PhD Thesis, Institute of Educational Technology, The Open University.

Jameson, A., Paris, C., and Tasso, C., editors (1997). *Proceedings of the Sixth International Conference on User Modeling in Sardinia, Italy*, Wien, New York. Springer.

Johnson-Laird, P. N. and Oatley, K. (1988). The language of emotions: An analysis of a semantic field. CSL Report 33, Princeton University. Cognitive Science Laboratory.

Kay, J., editor (1999). *Proceedings of the Seventh International Conference on User Modeling in Banff, Canada*, Wien, New York. Springer. Available on the World Wide Web at: `http://www.cs.usask.ca/UM99/` (May 24, 2000).

Keller, J. M. (1979). Motivation and instructional design: A theoretical perspectives. *Journal of Instructional Development*, **2**(4), 26–34.

Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth, editor, *Instructional-design Theories and Models: an Overview of their Current Status*, pages 383–434. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Kleinginna, P. R., J. and Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, **5**(4), 345–379.

Lang, P. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, **50**(5), 372–385.

Lazarus, R. S. (1991). *Emotion & Adaptation*. Oxford University Press, New York, NY.

Lepper, M., Woolverton, M., Mumme, D., and Gurtner, J. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. Lajoie and S. Derry, editors, *Computers as Cognitive Tools*, pages 75–105. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Lepper, M. R. and Chabay, R. W. (1988). Socializing the intelligent tutor: Bringing empathy to computer tutors. In H. Mandl and A. Lesgold, editors, *Learning Issues for Intelligent Tutoring Systems*, chapter 10, pages 242–257. Springer-Verlag, New York.

Lester, J. C., Converse, S. A., Stone, B. A., Kahler, S. E., and Barlow, S. T. (1997). Animated pedagogical agents and problem-solving effectiveness: A large-escale empirical evaluation. In du Boulay and Mizoguchi (1997), pages 23–30.

Lester, J. C., Towns, S. G., and FitzGerald, P. J. (1999). Achieving affective impact: Visual emotive communication in lifelike pedagogical agents. *International Journal of Artificial Intelligence in Education*, **10**, 278–291.

Lisetti, C. L. and Rumelhart, D. E. (1998). Facial expression recognition using a neural network. In *Proceedings of the 11th International Florida Artificial Intelligence Research Society Conference (FLAIRS'98)*, Menlo Park, CA. AAAI Press. Available on the World Wide Web at: `http://coba.usf.edu/departments/isds/faculty/lisetti/vita.htm` (October 1, 2001).

Malone, T. W. and Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In Snow and Farr (1987), chapter 10, pages 223–253.

Mandler, G. (1987). Emotion. In R. L. Gregory, editor, *The Oxford Companion to the Mind*, pages 219–220. Oxford University Press, Oxford.

Mark, M. A. and Greer, J. E. (1993). Evaluation methodologies for intelligent tutoring systems. *Journal of Artificial Intelligence in Education*, **4**(2/3), 129–153.

Matsubara, Y. and Nagamachi, M. (1996). Motivation system and human model for intelligent tutoring. In Frasson *et al.* (1996), pages 139–147.

Minsky, M. (1985). *The Society of Mind*. Simon & Schuster, New York, NY.

Murray, I. R. and Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal Acoustical Society of America*, **93**(2), 1097–1108.

Negroponte, N. (1996). *Being Digital*. Vintage Books, New York. Originally published by Alfred A. Knopf, Inc., New York, in 1995.

O'Bryen, P. (1996). Using questionnaires to assess motivation in second language classrooms. *University of Hawaií Working Papers in ESL*, **14**(2).

Ohlsson, S. (1987). Some principles of intelligent tutoring. In R. W. Lawler and M. Yazdani, editors, *Artificial Intelligence and Education. Volume One: Learning Environments & Tutoring Systems*, pages 203–237. Ablex Publishing, Norwood, N.J.

Oppenheim, A. N. (1992). *Questionnaire Design, Interviewing and Attitude Measurement*. Pinter Publishers, London and New York. First published as 'Questionnaire design and attitude measurement' by Basic Books, Inc., New York in 1966.

Ortony, A., Clore, G. L., and Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge.

Paiva, A., editor (2000). *Affective Interactions. Towards a New Generation of Computer Interfaces*, volume 1814 of *Lecture Notes in Computer Science*, Heidelberg. Springer-Verlag. Selected papers from (Affect in Interactions, 1999).

Parkinson, B. (1995). *Emotion*, chapter 1. In Parkinson and Colman (1995).

Parkinson, B. and Colman, A. M., editors (1995). *Emotion and Motivation*. Longman Essential Psychology. Longman, London.

Person, N., Klettke, B., Link, K., and Kreuz, R. (1999). The integration of affective responses into AutoTutor. In Affect in Interactions (1999).

Picard, R. W. (1995). Affective computing. Technical Report 321, M.I.T. Media Laboratory Perceptual Computing Section, Cambridge, Massachussetts. Available on the World Wide Web at: `http://vismod.www.media.mit.edu/cgi-bin/tr_pagemaker` (November 19, 1997).

Picard, R. W. (1997). *Affective Computing*. The MIT Press, Cambridge, Massachusetts.

Polson, M. C. and Richardson, J. J., editors (1988). *Foundations of Intelligent Tutoring Systems*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

Porayska-Pomsta, K., Mellish, C., and Pain, H. (2000). Aspects of speech act categorisation: Towards generating teachers' language. *International Journal of Artificial*

Intelligence in Education. Special Issue on Analysing Educational Dialogue Interaction, **11**(3), 254–272.

Reeves, B. and Nass, C. I. (1998). *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. Centre for the Study of Language and Information, US, New York.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., and Archer, D. (1977). The PONS test: Measuring sensitivity to nonverbal cues. In P. McReynolds, editor, *Advances in Psychological Assessment*. Jossey-Bass, San Francisco.

de Rosis, F. and Grasso, F. (1999). Affective natural language generation. In Paiva (2000), pages 204–218.

de Rosis, F., Grasso, F., and Berry, D. C. (1999). Refining instructional text generation after evaluation. *Artificial Intelligence in Medicine*, **17**(1), 1–36.

Roy, D. and Pentland, A. (1996). Automatic spoken affect analysis and classification. In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, pages 363–367, Killington, VT.

Scherer, K. R. (1981). Speech and emotional states. In J. K. Darby, editor, *Speech Evaluation in Psychiatry*, chapter 10, pages 189–220. Grune and Stratton, Inc.

Self, J. A. (1988a). Bypassing the intractable problem of student modelling. In *Proceedings of the First International Conference on Intelligent Tutoring Systems*, pages 18–24, Montreal, Canada. Sponsored by Universite de Montreal. In cooperation with Association for Computing Machinery.

Self, J. A. (1988b). Student models: What use are they? In P. Ercoli and R. Lewis, editors, *Proceedings of the IFIC TC3 Working Conference on Artificial Intelligence Tools in Education*, pages 73–86, Amsterdam. North-Holland.

Sloman, A. (1992). Prolegomena to a theory of communication and affect. In A. Ortony, J. Slack, and O. Stock, editors, *Communication from an Artificial Intelligence Perspective. Theoretical and Applied Issues*, volume 100 of *Series F: Computer and Systems Sciences*, pages 229–260. Springer-Verlag, Berlin. Heidelberg.

Sloman, A. (1999a). Contribution to 'Emotion' mailing list on 11th August 1999.

Sloman, A. (1999b). Review of *Affective Computing* by Rosalind Picard, 1997. *The AI Magazine*, **20**(1), 127–133.

Sloman, A. and Croucher, M. (1981). Why robots will have emotions. In *Proceedings of the 7th International Joint Conference on AI (IJCAI), Vancouver*, pages 197–202, San Mateo, CA. Morgan-Kaufmann.

Snow, R. E. and Farr, M. J., editors (1987). *Conative and Affective Process Analyses*, volume 3 of *Aptitude, Learning, and Instruction*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

Sobral, D. (1992). Self-report visual scale of course appeal. *Higher Education*, **23**, 321–329.

Socially Intelligent Agents (2000). *Proceedings from the Symposium 'Socially Intelligent Agents - The Human in the Loop'. Technical Report FS-00-04*. AAAI Press.

del Soldato, T. (1994). *Motivation in Tutoring Systems*. Ph.D. thesis, School of Cognitive and Computing Sciences, The University of Sussex, UK. Available as Technical Report CSRP 303.

Spensley, F., Elsom-Cook, M., Byerley, P., Brooks, P., Federici, M., and Scaroni, C. (1990). Using multiple teaching strategies in an ITS. In C. Frasson and G. Gauthier, editors, *Intelligent Tutoring Systems: At the Crossroads of Artificial Intelligence and Education*, pages 188–205. Ablex Publishing Corporation, New Jersey.

Spitzer, D. R. (1995). *SuperMotivation*. AMACOM Books, New York.

Spitzer, D. R. (1996). Motivation: The neglected factor in instructional design. *Educational Technology*, **36**(3), 45–49.

Stoney, S. and Oliver, R. (1998). Interactive multimedia for adult learners: Can learning be fun? *Journal of Interactive Learning Research*, **9**(1), 55–82.

Tosa, N. and Nakatsu, R. (1996). Life-like communication agent–emotion sensing character 'MIC' and feeling session character 'MUSE'. In *Proceedings of the International Conference on Multimedia Computing and Systems*, pages 12–19. IEEE.

Turban, E. (1992). *Expert Systems and Applied Artificial Intelligence*. Macmillan Publishing Company, New York.

VanLehn, K. (1988). Student modeling. In Polson and Richardson (1988), chapter 3, pages 55–78.

Vedder, P., editor (1990). *Fundamental Studies in Educational Research*. Swets & Zeitlinger, Amsterdam.

de Vicente, A. and Pain, H. (1998). Motivation diagnosis in intelligent tutoring systems. In Goettl *et al.* (1998), pages 86–95.

Vidler, D. C. (1977). Achievement motivation. In Ball (1977), chapter 4, pages 67–89.

Warwick, K. (1998). *March of the Machines. In the Mind of the Machine: the Breakthrough in Artificial Intelligence*. Arrow, London.

Weiner, B. (1984). Principles for a theory of student motivation and their application within an attributional framework. In Ames and Ames (1984), chapter 1, pages 15–38.

Weiner, B. (1992). *Human Motivation: Metaphors, Theories, and Research*. Sage Publications Inc., Newbury Park, London.

Weinert, F. E. (1990). Theory building in the domain of motivation and learning in school. In Vedder (1990), chapter 5, pages 91–120.

Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge*. Morgan Kaufmann Publishers, Inc., Los Altos, Calif.

Whitby, B. (1996). *Reflections on Artificial Intelligence. The Legal, Moral and Ethical Dimensions*. Intellect Books, Exeter, Great Britain.

Whitelock, D. and Scanlon, E. (1996). Motivation, media and motion: Reviewing a computer supported collaborative learning experience. In Brna *et al.* (1996), pages 276–283.

Williams, M. and Burden, R. L. (1997a). *Psychology for Language Teachers: A Social Constructivist Approach*. Cambridge University Press, Cambridge, UK.

Williams, M. and Burden, R. L. (1997b). What makes a person want to learn? Motivation in language learning. In Williams and Burden (1997a), chapter 6, pages 111–142.

Wlodkowski, R. J. (1998). *Enhancing Adult Motivation to Learn*. Jossey-Bass, San Francisco.

Yacoob, Y. and Davis, L. S. (1996). Recognizing human facial expressions from log image sequences using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(6), 636–642.

Yin, J., El-Nasr, M. S., Yang, L., and Yen, J. (1998). Incorporating personality into a multi-agent intelligent system for training teachers. In Goettl *et al.* (1998), pages 484–493.

Appendix A

Self-report study materials

This appendix contains the following materials in relation to the self-report study: a description of the prototype ITS used, the instructions given to the participants, and the post-questionnaire administered to participants.

A.1 Prototype ITS 210

A.2 Materials given to students prior to interaction 216

A.3 Post-questionnaire given to participants 220

A.1 Prototype ITS

In order to perform the self-report study described in chapter 4, we developed a simple tutoring system with two main added features:

1. It allows the student to inform the system about his motivational state as the instruction takes place. This feature was added to test the feasibility of self-report as a motivation diagnosis technique.

2. Student interactions can be recorded. These interactions can be replayed at a later stage. This feature was added to perform a Motivation Diagnosis study in which teachers' motivation diagnosis techniques could be elicited (details of this are given in chapter 5).

A.1.1 Instructional domain

The domain chosen for the prototype ITS was numbers in Japanese[1], as these are sufficiently difficult that they need many hours of practice until a student masters them, but at the same time a system to teach them can be easily implemented. To avoid making the task unnecessarily difficult, numbers were not taught using the actual Japanese syllabary. Instead, we used the *romaji* syllabary, which is a transcription into latin alphabet of the Japanese *kana*, and which is widely used to teach Japanese to westerners.

Five types of lessons were developed: 1) simple presentation of numbers to memorise; 2) identification of numbers written in Japanese from a set of four possible answers; 3) identification of numbers from a set of four possible answers, but presenting the numbers in context (either as time or as prices of certain goods); 4) writing the Japanese numbers given the numeral; and 5) game with numbers. For each of the first four types we developed 2 lessons, one dealing with numbers up to 20 and another dealing with numbers up to 100. For the game we developed just one lesson dealing with numbers up to 20, as the difficulty was already considerable. Table A.1[2] gives the description of the implemented lessons and figure A.1 presents the main interface.

[1] We thank Madoka Aiki for checking the correctness of the teaching materials.

[2] This table appears in the main text of this dissertation as table 4.2

Type		Lesson	Description
1.	Theory	Theory20	Explanation of Japanese number system (up to 20)
		Theory100	Explanation of Japanese number system (up to 100)
2.	Identification	Identi20	Identifying Japanese numbers (up to 20)
		Identi100	Identifying Japanese numbers (up to 100)
3.	Identification	Context20	Identifying time (up to 20)
	in context	Context100	Identifying prices (up to 100)
4.	Writing	Writing20	Writing Japanese numbers (up to 20)
		Writing100	Writing Japanese numbers (up to 100)
5.	Game	Game20	Game with numbers (up to 20)

Table A.1: Description of instructional units

The task to be performed in each of the lessons is straightforward to understand by the student and easily implemented, except for lesson Game20, which consists of the educational game described below.

Numbers game[3]

A screen-shot of the game interface can be seen in figure A.2. A simple arithmetic expression (in figure A.2 *yon* + *ju-kyu*, i.e. 4 + 19) falls through a passage, the student has to read the numbers in Japanese, calculate the arithmetic result and select the appropriate answer from a set of given possibilities. In figure A.2 there are eight possibilities given to the student (these are greyed out simply because the game was paused), the correct one being *ni-ju-san*, i.e. 23.

If the choice made by the student is the correct one, the score will increment by 5, the arithmetic expression will be deleted from the passage and a new one will start falling from the top of the passage. The student can select an incorrect answer only twice, when the arithmetic expression will fall into the bottom of the passage, making it smaller, and therefore making the next exercise more difficult. As another form of

[3]We thank Jeff Hobbs for providing the original Tetris game source code, on which this game is based. Jeff's code is available at the World Wide Web: `http://sunscript.sun.com/plugin/tetris.html` (September 8, 1998).

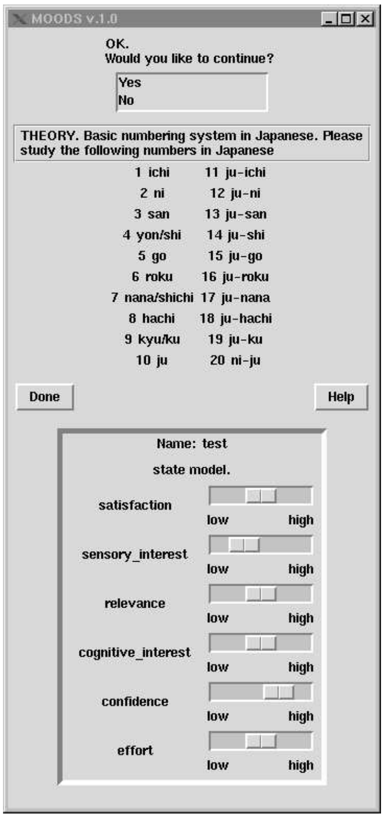


Figure A.1: Main interface

incentive to give correct answers, the speed on which the expressions fall will be increased every time the student makes a wrong choice or when the expression is allowed to reach the bottom of the passage.

As time passes by, and to avoid continuing in the same level of difficulty for long, the speed increases steadily (every time that 20 points are obtained, except at the beginning of the game, when it remains at speed 1 until the first 40 points are obtained). Similarly, the number of possible answers shown to the student varies from 6 when the speed is low to 10 options when the speed is high, making the game increasingly difficult to play. If not finished earlier, the game will finish when the speed level is 40 or higher, which is too fast for practical playing in the machines available for our project[4].

[4] ULTRA 5 Sun Workstations

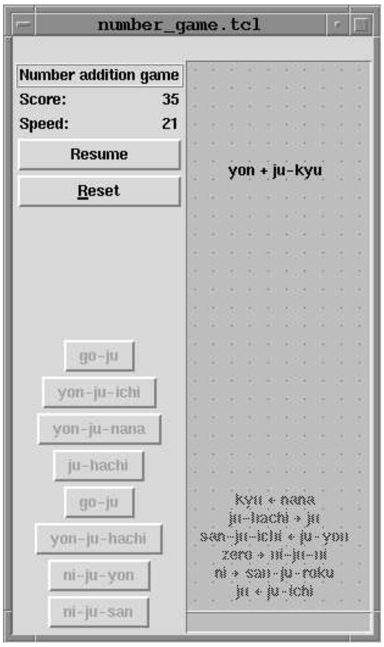


Figure A.2: Numbers game

But normally the game will end when, due to student's mistakes, the passage has been made very small, and there is no space for further arithmetic expressions. The score will tell us how well the student performed.

A.1.2 Usage description

On its first use, the system will ask the user about his name and about the trait characteristics explained in chapter 3. The questionnaire presented to the student is given in figure A.3.

After this has been filled in, the main interface will appear, which we have already seen in figure A.1. The interface consists of three different parts. On the top side, we have the dialogue frame, which is only visible when the computer communicates with the student (eg. to give feedback). In the middle we have the lessons, and on the bottom part we have the sliders that represent the motivational state, which the user is free to update as the instruction takes place.

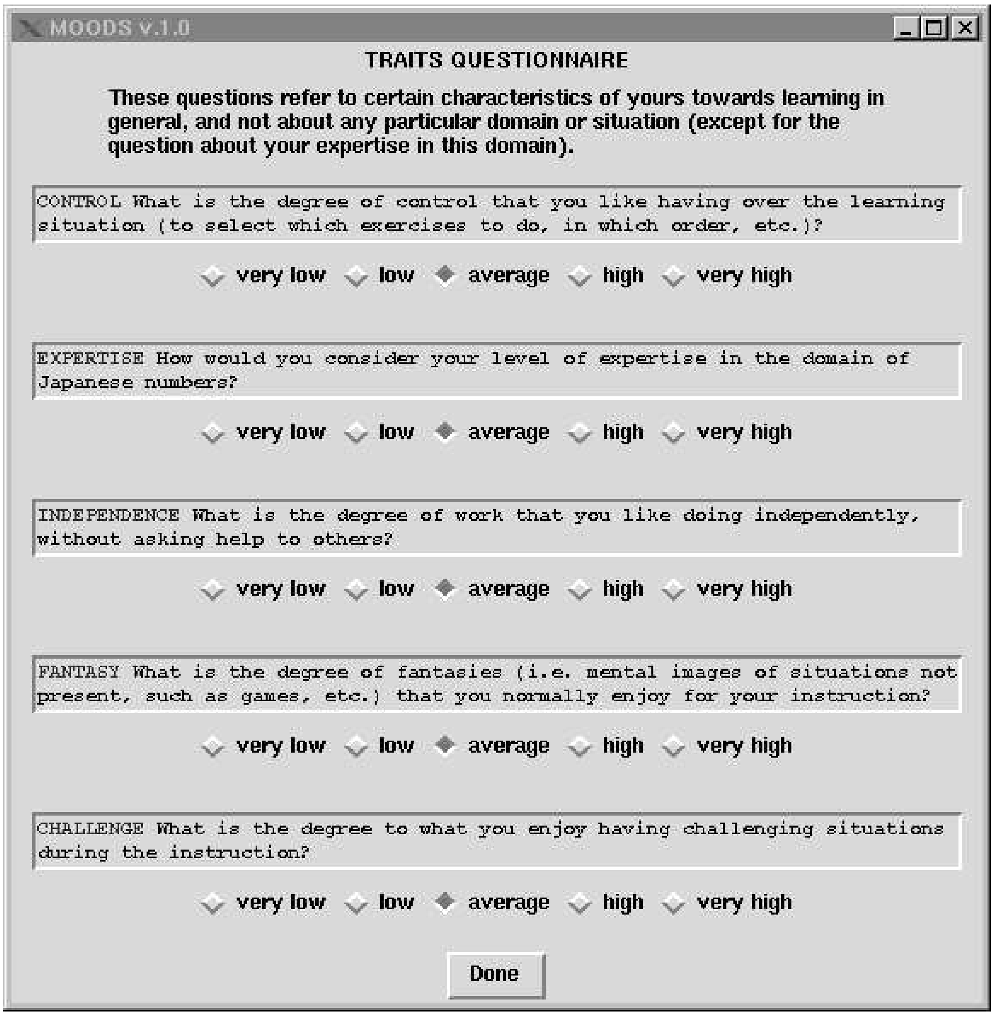


Figure A.3: Trait characteristics questionnaire

Feedback

The feedback given to the student is very simple. The system presents a message to the student whenever he clicks in any one of the buttons provided (‘done’, ‘give’ or ‘help’). The help messages relate to the particular instructional unit active at the moment, and the number and content of messages for different lessons varies from 0 to 3 in the developed prototype version.

The feedback given after pressing the buttons ‘done’ or ‘give up’ depends solely on how well the student was performing. This is summarised in tables A.2 and A.3.

One of two measures is consulted, Mark or Mark_done. Mark is a total outcome while Mark_done takes into consideration how many questions the student answered. For example, if there are 10 questions, the student answers only 4 and from these four gets correct 3, Mark will be 30% while Mark_done will be 75% (3 out of 4). This is useful for the feedback to the Give_up button. If the student didn't try very hard, but, as in this example, got many of the answered ones correct, then the feedback should be encouraging. In all these cases, the message will be followed by a message asking whether the student would like to continue with another instructional unit or not.

Mark	**Feedback**
0	Sorry, you didn't get any!
≤ 30	Uhhm, I think you should practice a bit more.
≤ 60	OK, not too bad.
≤ 80	That was pretty good.
≤ 100	Well done!

Table A.2: DONE feedback

Mark_done	**Feedback**
0	Well, but you didn't try very hard, did you?
≤ 50	Uhhm, I see. Maybe you want something easier.
≤ 80	You want to give up? You were not doing too badly!
≤ 100	Well, you were doing really well!

Table A.3: GIVE UP feedback

A.2 Materials given to students prior to interaction.

Instructions for experimental study

Thanks for agreeing to participate in this study. Your collaboration will be used as part of ongoing research on the design of more efficient computer-assisted instructional systems. The study will take approximately 30 minutes, in which you will be asked to use a prototype of an instructional system, which is described below.

The system you will use is a prototype of a tutor of Japanese numbers and you don't need to have any prior knowledge of Japanese.

The interaction will be as follows: when you type in your name, a very small questionnaire will pop up, which will ask you for your preference for five factors that the system will take into consideration when delivering the instruction. When you have answered this questionnaire, you will see the main interface of the system, similar to that in the figure given.

The upper part of the interface contains the instructional unit. The units are mostly straightforward exercises. You will have two or three buttons to indicate when you have finished, or whether you want to give up or ask for help. Use these whenever you think you need them.

On the lower part of the interface you can see six sliders that you can use to communicate to the computer your current state at any given moment (for instance about your confidence in solving the problem, about the interest that a given exercise arouses in you, etc.). Please, use these sliders **as often as possible** whenever you think there is a change in any of these factors, since it is necessary for the computer to understand your current situation in order to modify the instruction accordingly. A brief explanation of what is meant by each of these factors is given to you as a separate sheet.

Remember that you are not supposed to cover any particular set of materials, but rather spend some time with the system and try to learn the basics of the Japanese counting system. There is no good or bad interaction with the system and everything you do will only be used anonymously to publish results about the experiment.

Thanks,

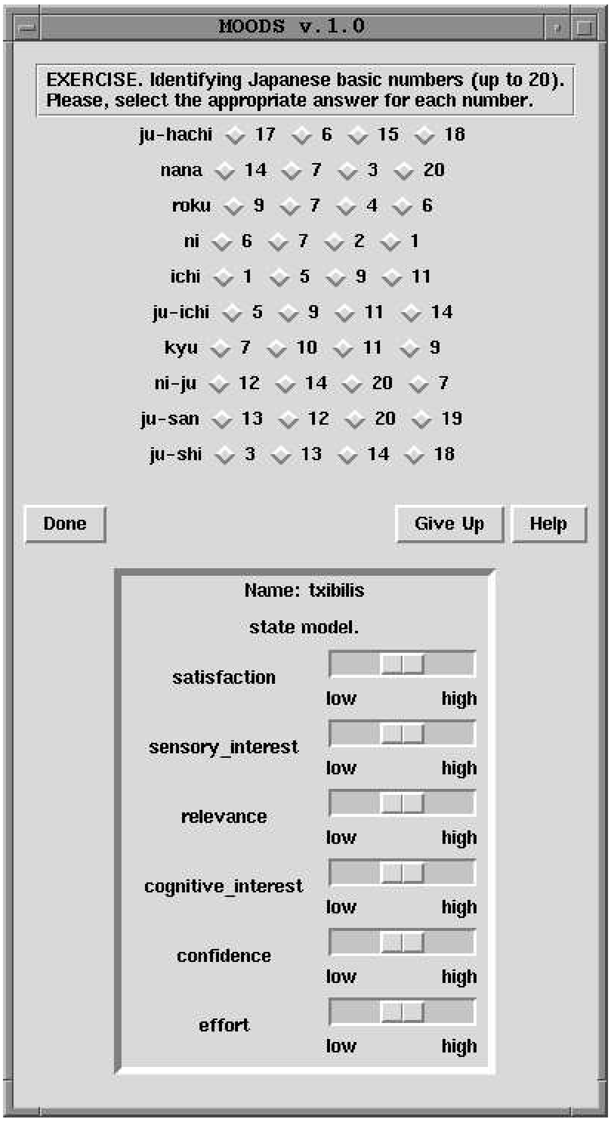


Figure A.4: Main user interface

Definitions

Understanding these definitions is important for this study. Please make sure that you understand them, and please, don't hesitate to ask for clarification if needed.

- Trait characteristics (*These represent some more or less 'stable' characteristics of yours (i.e. that don't change much over time). The computer will ask you about these characteristics at the beginning of the interaction in a small questionnaire.*)
  - **Control** What is the degree of control that you like having over the learning situation? (i.e. do you like to select yourself which exercises to do, in which order, etc. rather than let the instructor take these decisions)?
  - **Independence** What is the degree of work that you like doing independently, without asking others for help? (i.e. do you prefer to work on your own, even if you find some difficulties, and try to solve them by yourself rather than asking collaboration or help from others?)
  - **Fantasy** Would you enjoy having fantasies (mental images of situations not present, such as games, etc.) in your instruction? (for instance, learning addition can be done through simple theory or through a 'fantasy' game in which a rocket advances through space every time your addition is correct. Would you like that?)
  - **Challenge** What is the degree to which you enjoy having challenging situations during the instruction? (i.e. would you like exercises that go a little bit beyond your current level of understanding and that represent a challenge to you?)
  - **Expertise** How would you consider your level of expertise in the domain of Japanese numbers?
- State factors (*These represent transient factors of yours (i.e. factors that are likely to change during the course of the instruction). These will be represented in the main interface as sliders that you will have the chance to update regularly.*)

- **Satisfaction** refers to the overall feeling of goal accomplishment. (i.e. do you think that the instruction is satisfying and that is making you closer to your goals?)
- **Relevance** refers to personal needs. A situation will be seen as relevant if the learner perceives that *important personal needs* are being met by the learning situation. (i.e. Do you think that the instruction is personally relevant to you?)
- **Confidence** refers to the student's belief in being able to perform the task at hand correctly. (i.e. Do you think you will be able to do the exercise shown to you correctly?)
- **Interface interest (or sensory interest)** refers to the amount of curiosity aroused through the interface presentation. (i.e. Do you find the graphics, sounds, etc. appealing?)
- **Task interest (or cognitive interest)** refers to curiosity aroused through the cognitive or epistemic characteristics of the task. (i.e. regardless of the presentation issues, do you think the task at hand is cognitively appealing?)
- **Effort** refers to the amount of work that the student is doing. (i.e. How much work would you say you are doing in order to solve the problem given?)

A.3 Post-questionnaire given to participants

Questionnaire

Now that you have finished the experiment, we would like to ask you to fill in the following questionnaire, which will help us to understand your interaction with the system better. If in doubt, please feel free to ask for further explanation, and please, if you feel like it, do add any extra comments.

Thanks.

	1(strongly agree) . . . 5(strongly disagree)
About the experiment	
1) My role in this experiment was clear	1 2 3 4 5
2) The time needed to perform the experiment was too long	1 2 3 4 5
General impressions about the system	
3) The system was easy to use	1 2 3 4 5
4) I think the system could be useful for learning Japanese numbers	1 2 3 4 5
5) The system seemed to react appropriately when I updated the motivational factors sliders	1 2 3 4 5

About the trait characteristics questionnaire (Control, Independence, Challenge, Fantasy, Expertise)

6) The definitions of the trait factors were easy to understand 1 2 3 4 5

*If you **disagree**, could you tell which ones were difficult to understand?*

7) The questions about trait characteristics were easy to answer (I easily know my 'level' for each of those characteristics) 1 2 3 4 5

*If you **disagree**, could you tell which ones were difficult to answer?*

8) I don't think that there are any other trait characteristics that it would be useful for the system to know about 1 2 3 4 5

*If you **disagree**, could you tell which characteristics about yourself you think could be useful for the system to know?*

9) I would prefer not to answer the trait questionnaire, even if it makes the instruction more efficient and personalized 1 2 3 4 5

About state factors (Satisfaction, Relevance, Sensory interest, Cognitive interest, Confidence, Effort)

10) The definitions of the motivational state factors were easy to understand 1 2 3 4 5

*If you **disagree**, could you tell which ones were difficult to understand?*

11) The sliders representing motivational state factors were easy to answer (I easily know my 'level' for each of those factors) 1 2 3 4 5

*If you **disagree**, could you tell which ones were difficult to answer?*

12) I don't think that there are any other state factors that it would be useful for the system to know about 1 2 3 4 5

*If you **disagree**, could you tell which characteristics about yourself you think could be useful for the system to know?*

13) I would prefer not to have to update the motivational state sliders, even if it makes the instruction more efficient and personalized 1 2 3 4 5

Do you have any extra comments?

Appendix B

Motivation Diagnosis study materials

This appendix contains the following materials in relation to the Motivation Diagnosis study: the instructions given to the participants, a transcription of the interview with one of the participants, and the complete set of the first version of the motivation diagnosis rules inferred from the study.

B.1 Instructions to participants 224

B.2 Transcription of interview 232

B.3 First version of Motivation Diagnosis Rules 254

B.1 Instructions for participants.

Introduction

First of all, thanks for being willing to participate in this study. Your collaboration will take approximately 1 hour and it will be used as part of ongoing research concerning the design of more efficient computer-assisted instructional systems.

Briefly, the goal of this study is to explore issues of diagnosis of students' motivation during instructional interactions. The task we ask you to perform for this study involves watching recorded interactions of a student with an instructional system, and inferring and commenting on the affective state of the student at certain points during the instructional interaction.

In order to perform this study, you will use a computer system (whose functioning is explained below) and you will be encouraged to verbalise the reasoning behind your inferences about the affective state of the student. I will be present during your collaboration in order to prepare the equipment, solve any possible problems with the software, and help you to verbalise your inferences. Your comments and your interaction with the computer system will be recorded, but they will be used strictly anonymously for the analysis of data.

In the sections below I explain your task in greater detail. Please read through them carefully, and if you have any doubts please do not hesitate to ask for clarification at any point.

Interface for the study

The interface you will use consists of two windows and can be seen in figure B.1.

1. You can think of the window to the left (MOODS v.1.0) as a kind of video recorder. In it you will be able to see replayed the interaction of a student with a tutoring system.

2. The window to the right (Motivation model) consists of three frames.

   (a) The top frame is a representation of a number of *student traits*, which provides information about some general learning characteristics of the student.

   (b) The middle frame contains three buttons that allow you to control how the student interaction is replayed.

   (c) The bottom frame is a representation of the student's *motivational state*, which contains a number of sliders that represent certain motivational variables of the student. Your task involves predicting the likely values of these motivational variables as the instruction takes place. Therefore, it is only this frame that you will have to manipulate.

   The rest of the system interface gives you information about the student and the instructional interaction, which you will need in order to make inferences about the student's motivational state.

In the following sections we discuss in greater detail each of these parts of the system.

Instructional interaction

As mentioned above, the student's interactions are replayed in the left window in figure B.1. These interactions have been recorded previously as students used MOODS, a simple tutoring system to learn Japanese numbers. None of the students had any knowledge of Japanese before the interaction with MOODS took place.

To avoid making the use of MOODS unnecessarily difficult, numbers were taught using the *romaji* syllabary, which is a transcription into Latin alphabet of the Japanese *kana*, and which is widely used to teach Japanese to westerners. MOODS was implemented with five types of lessons:

1. Theory: simple presentation of numbers to memorise;
2. Practice: identification of numbers written in Japanese from a set of four possible answers;
3. Practice: identification of numbers from a set of four possible answers, but presenting the numbers in context (either as time or as prices of certain goods);
4. Practice: writing the Japanese numbers given the numeral; and
5. Game: playful activity based on the well-known Tetris game.

The interaction you will see is made up of a combination of lessons of these types, although not necessarily all of them. In figure B.1 you can see an example of a lesson of type 2. An example of a lesson of type 3 can be seen in figure B.2.

Although MOODS teaches Japanese numbers, you need no prior knowledge of Japanese to perform this study. The program originally used by students has been modified so that the correct solution to the exercises is shown between brackets. We can see an example of this in both figures B.1 and B.2. At the same time, in order to follow more easily the interaction with MOODS, an arrow (as seen in figure B.1) indicates the mouse movements performed by the students.

Controlling the replay of the interaction

The three buttons in the middle frame of the Motivational Model window in figure B.1 let you control the interaction at will, which is guided by the following principles:

- The replay of the student interaction will start when you press "Play"and stop when you press "Stop".
- The button "Initialise Lesson" will allow you to rewind the interaction to the beginning of the current lesson, in case you would like to see a part of the interaction again.
- The replay of the interaction is done in real time, except for the replay of the theory lessons (i.e. lessons of type 1). In this type of lesson there is very little student activity, and therefore the system will simply show a message informing you of how long the student took to learn the lesson (see figure B.3), and afterwards the interaction continues.
- In order to let you update the student's motivational model (discussed in the next section), the interaction will stop by default in the following three cases:
  1. When the student presses any of the buttons (Done, Give Up, or Help), but before any feedback is given by the system. At this point we would like you to verbally comment on the type of feedback that you think would be the most appropriate to give to the student.
  2. When feedback is presented to the student.
  3. When a new lesson is presented to the student.

  In these cases a small message similar to that in figure B.4 will be shown to remind you of updating the motivational model. But if you would like to update the motivational model or comment on any aspect of the instruction at any other time, you can, as mentioned above, stop the interaction by pressing the "Stop" button.

The Motivational Model of the student

The Motivational Model window in figure B.1 contains a representation of the student's motivational model, which is given by two differentiated parts:

1. Student traits (top frame of window). These represent permanent characteristics of the student, and are given solely to help you infer how different aspects of the interaction influence the student's motivational state.

2. Student motivational state (bottom frame of window). This is a representation of a number of affective factors of the student whose values during the interaction you are asked to infer and comment upon.

The definition of these characteristics and factors is given below. Please read them carefully before continuing.

- Trait characteristics: *These represent stable characteristics of the student, i.e. are not likely to change during the instructional interaction.*

  - **Control** refers to the degree of control that the student likes having over the learning situation (i.e. does he[1] like to select which lessons to study, in what order, etc. himself, rather than let the instructor take these decisions?).
  - **Fantasy** refers to whether the student would enjoy having fantasy environments (mental images of situations not present, such as games, etc.) during the instruction (for example, whether he would like learning addition through a 'fantasy' game in which a rocket advances through space every time that his answers are correct).
  - **Independence** refers to the degree of work that the student likes doing independently, without asking others for help (i.e. does he prefer to work on his own, and if he finds some difficulties try to solve them by himself, rather than asking for collaboration or help from others?)

[1] In order to increase the clarity of these instructions we use the masculine pronoun to refer either to he or she.

  - **Challenge** refers to the degree to which the student enjoys being challenged during the instruction (i.e. whether he would like lessons that go a little bit beyond his current level of understanding and that represent a challenge to him).

- State factors: *These represent transient characteristics of the student (i.e. characteristics that are likely to change during the course of the instruction). These are represented in the interface as sliders that you will have the chance to update regularly.*

  - **Satisfaction** refers to the overall feeling of goal accomplishment (i.e. whether the student perceives that the instruction is satisfying and whether it is moving him closer to his goals).
  - **Sensory interest (or interface interest)** refers to the amount of curiosity aroused through the interface presentation (i.e. whether the student finds the graphics, sounds, etc. appealing).
  - **Cognitive interest (or task interest)** refers to curiosity aroused through the cognitive or epistemic characteristics of the task (i.e. regardless of the presentation issues, whether the task at hand is perceived by the student as cognitively appealing).
  - **Confidence** refers to the student's belief in being able to solve the lesson at hand correctly.
  - **Effort** refers to the amount of work that the student is doing.

Given these definitions, an example inference could be made in a situation when the student gives up on a lesson after a very short time and without attempting to give answers for any of the exercises, in which case you could perhaps infer that his overall satisfaction and his effort are very low at this time during the interaction.

Summary of your task

To summarise, your task will consist of the following:

1. You will be presented with information about certain trait characteristics of a student (who had no knowledge of Japanese before the interaction with MOODS took place).

2. You will be shown a replay of his interaction with the MOODS instructional system.

3. Throughout the interaction, and particularly at any stop points, you are encouraged to give verbal comments on the student's motivational state and the possible factors affecting it. *For ease of analysis, we will be taping your comments. Please indicate now if you have any objections to this.*

4. Whenever the interaction is paused, you are asked to update the motivational state variables. But you only need to update those variables for which you think you have enough information to create an inference. As stated in point 3 above, you should attempt to verbalise the reasoning behind your inferences in as concrete terms as possible: I may ask you for further clarification.

5. When the student presses any of the three buttons available to him (Done, Give up, or Help) you will also be encouraged to comment on the type of feedback that you think would be the most appropriate to give to the student.

Before you start, just to remind you once more that if you have any questions, please do not hesitate to ask.

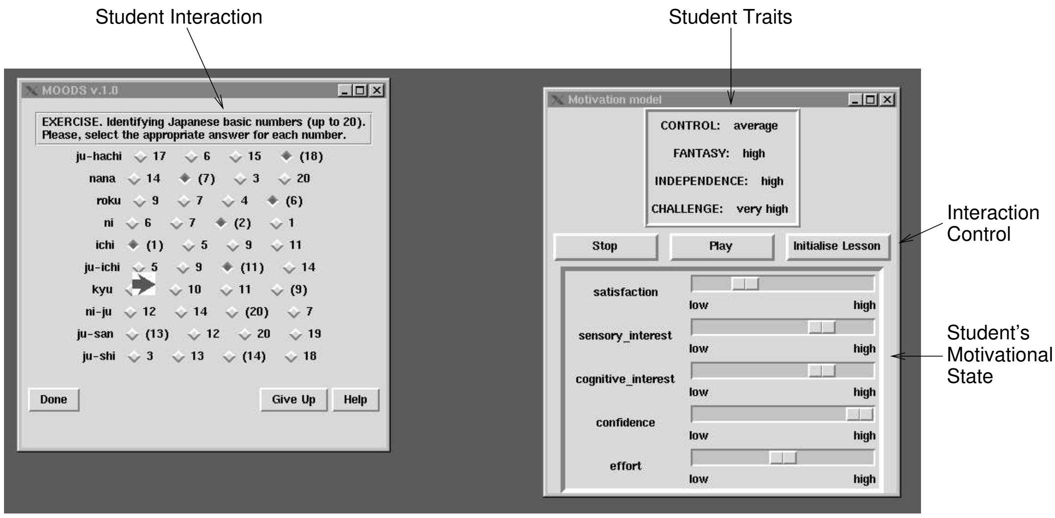


Figure B.1: Interface for study.

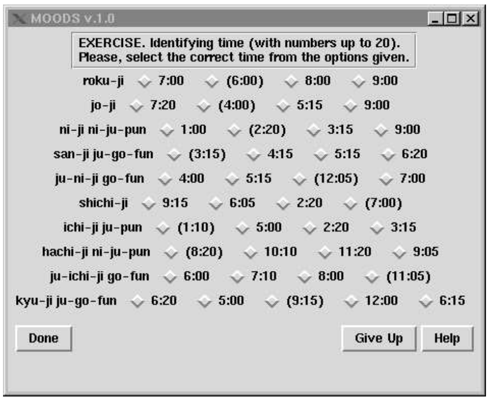


Figure B.2: Example of lesson.

Figure B.3: Sample message to inform of time spent in theory lesson.

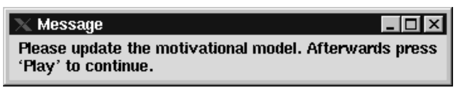


Figure B.4: Sample message to remind of updating model.

B.2 Complete transcription of one interview

The following transcription corresponds to the interview with participant number 4. For ease of reference, we have highlighted the text in the transcription that gave rise to the motivation diagnosis rules of participant 4, indicating which rule the text gave rise to. The actual rules for participant number 4 can be seen in table B.5.

Some introductory and concluding comments during the interview are not included in this transcription, as they were mainly remarks about the study settings and/or social conversation, which are not relevant for the study. We have attempted to transcribe the interview as truthfully as possible, and therefore a number of grammar mistakes can be found in it. Between brackets we can find the time during the interview (in minutes and seconds) when a particular remark was made.

(3.22) Interviewer: So, this task many people find it quite difficult to do, so if sometimes you don't have anything, because the interface is not very rich, in the type of things that you can see from the student, if you cannot infer anything, you can just continue, but sometimes the interface will actually give you more information. In those cases you can comment something. So if we start recording, I press the first Play for you, and then now

(3.50) Participant: So I have control, I can stop it if I want?

(3.50) Int: Yes. So at the beginning this one is just theory, so not to stay too much in this one, you know more or less how much time he spent on this lesson, spent like 102 seconds, around 2 minutes, just trying to memorise these numbers in Japanese. So, the interaction will happen always the same, it will stop when he clicks done or any of the buttons, so if you think that by the amount of time that he spent or something similar you can, and by the traits that he said about himself, if you think that you can say something about those ones, and basically, what I would like is, if you move anything try to verbalise in terms of what you saw in the interaction and in terms of his own traits.

(4.49) Par: OK, so I think he is quite satisfied with that because he did it quite quickly, and I don't have any comments about those. It is hard to tell if he was cognitively interested or not.

(5.03) Int: Yes, most of the time will be quite hard.

Rule (a)

(5.10) Par: And I'd say he was quite confident, well reasonably confident.

(5.10) Int: And that based on?

(5.14) Par: Because again, he did it quite quickly, even though it is quite hard to tell because you can't see, but he seemed confident, because it seemed that he looked at it quickly and then pressed done.

(5.30) Int: So, it was basically 20 items in around 2 minutes.

(5.42) Par: These ones again, how do I know, ah, this is what is telling me about the student, isn't it?

(5.46) Int: Yes, so this is how the student characterises himself, in terms that he enjoys challenging situations.

(5.54) Par: I will have to leave effort in the middle. I don't know what effort he made.

(6.00) Int: Yes, at the beginning, until there is a bit more of information, it is difficult.

(6.03) Par: OK, so I go for the next one.

(6.06) Int: So again, now the student is given some type of feedback, sometimes will be difficult to update anything, sometimes it will be a bit easier, so from this feedback, do you think any of those factors could change, then you can infer something, otherwise you can continue. So, basically the student clicked "Done" and the system tells "OK, would you like to continue?"

(6.32) Par: So, this is right after he pressed done?

(6.34) Int: Yes.

(6.36) Par: So I don't know if he is going to say Yes or No?

(6.38) Int: No, not for the moment. Because the system stops always in those three places, so now is the same stuff, he pressed that he would like to continue and another lesson, in this case another theory lesson is shown, so again, maybe from the type of lesson that is given to him, I don't know if you could ...

Rule (b)

(7.04) Par: Yes, so I think is basically the same, as it was when he first looked at the theory, I think he is still quite interested [noise] he is satisfied, so I think this is the same, maybe I would push this one up a bit, cognitive interest.

(7.16) Int: And why?

(7.19) Par: Because he is continuing with the lesson.

(7.24) Int: Yes, I mean, in the theory ones is difficult to tell. So in this one he spent 142 seconds, so a bit longer than in the previous one, above two minutes, and now he pressed "Done"

(7.50) Par: What was sensory interest again?

(7.54) Int: Sensory interest refers to whether the interface itself, the graphics, the sound, which there is not here, actually arouses the student, creates any interest.

(8.03) Par: Well, I don’t think it does, I think he is not spending enough time on it. I learnt Japanese before, does that affect us?

(8.14) Int: No, I mean, basically you will be able to understand better what he is doing.

(8.15) Par: Well, I don’t think he is spending enough time on it. Is he a beginner?

(8.20) Int: Yes, all these people had no idea whatsoever [of Japanese].

(8.22) Par: He is not aroused by it at all, I shouldn’t think so.

(8.26) Int: So, but the sensory interest refers only to the graphics, sounds, etc. and the task itself is the cognitive interest. So, and these ones you are changing....

Rule (c)

(8.40) Par: Yes, I'm changing these down because there is an awful lot of information there, and if he is a beginner, if it only took him 142 seconds, and he is moving on, is the idea that he looks at these and then he moves on to do some exercises or something, so that is not enough time, so I don't personally think this is good ... graphic.

(9.00) Int: So, then you think because he didn't spend too much time on it, then and he probably should spend more because the task is difficult, then that means that his cognitive interest is probably going down.

(9.14) Par: Yes, if he was going to read out each of these numbers himself, and practice speaking and pronouncing them sure it would take longer than 142 seconds.

(9.24) Int: I agree, I've done it myself and it takes long...

(9.30) Par: But he might be satisfied with this, but I don't know, I cannot answer this, so I will leave it in the middle.

(9.33) Int: Later will become a bit easier when he starts doing the exercises.

(9.38) Par: OK, next one.

(9.38) Int: So, again, this is plain feedback.

(9.43) Par: I can't. I don't think I will change anything.

(9.50) Int: So now, this is the first exercise that he has to do. This is the one that I showed to you before. He has these numbers here, and these four options, and he will have to choose, so ...

(10.08) Par: He has to go through all the lines?

(10.10) Int: Well, yes, basically the exercise is to go through all of them.

(11.28) Int: OK, so now, basically, now you see before the system tells the student how many he got correct and what type of feedback it is going to give, then now the same as we said before, well first of all, if you can comment on what type, knowing which ones he got right, and which ones he got wrong, etc., if you could comment on the type of feedback you think it would be appropriate to give, and then if you can infer something about his motivational state at this point, based on his traits and based on what you just saw in the interaction.

(12.01) Par: OK, so

(12.02) Int: If you want you can move this window in order to see the..

(12.11) Par: Well, he did really well, he only got one wrong, he didn't answer this one, and he got this one wrong, and this twelve and twenty is easy to mix up in Japanese, so I don't know if he needed any feedback, I think the exercise it was very clear what he had to do and I think getting feedback after he's answered is better, you know proper feedback, I don't know, he hasn't got any feedback yet, has he?

(12.46) Int: No, not yet

(12.48) Par: He will?

(12.50) Int: Basically the system stops now and he will get some feedback, but now you have the opportunity to comment on the type of feedback that you would give him.

(12.57) Par: Ah, OK, I think the ones he got right they should maybe go through each one of them individually and say, OK so this one is 18, ju-hachi, correct, maybe each one, and the ones he got wrong maybe he should do again, have another, the one he got wrong and the one he left out, he should have another chance, maybe?

(13.30) Int: Obviously, there is no right or wrong answer, and it is actually very difficult but... and about the factors, do you think you could infer anything?

(13.41) Par: I think he is very confident, not very confident, quite confident, he did move about a bit, but I think he was doing, looking at what he had, he got this, when he clicked 13 it was very easy for him to click 14, so, you know because the ju-san, the ju-shi, were quite similar to each other, so it was kind of, in a way, he is maybe he shouldn't, maybe they shouldn't be in a list like that, because having this one helped him quite quickly to get the one underneath it, so maybe if each row was shown to him first and then another row, and another row, not in a line like this.

(14.27) Int: But then, you said that you think that he looks quite confident, I mean, how can you know that from what you saw in here?

(14.36) Par: I said that he was quite confident because, because he got them right?

(14.45) Int: But he doesn’t know that yet.

(14.50) Par: Well, I can’t really say that.

(14.53) Int: You said something before about that he didn’t move too much or ...

Rule (d)

(14.59) Par: He did move a bit at the beginning, but I think he was looking at everything that was in front of him, he was taking it all in, he wasn’t just going through one line at a time, like this, if he was very confident he would just do that and click very quickly so I’m going to move that down, but he took maybe I don’t know, he took maybe three rows at a time and looked at all the numbers and the possible answers, use the other possibilities from the other questions to help him answer the one that he was trying to answer at the time, so maybe, I don’t know if that is such a good idea. Maybe he should get just one row at a time, you know what I mean?

(15.45) Par: I can’t tell if he was interested. Maybe this is just really boring for him.

(15.50) Int: Yes, maybe sometimes you don’t have enough information, I mean you also have to take into account what he said about himself, perhaps that can tell you something.

Rule (e)

(16.02) Par: Well, he likes a lot of fantasy, so I doubt that he is sensory stimulated at all, I'm going to put that lower. Because if he likes a lot of fantasy, this should be colourful and, you know.

(16.17) Int: And about the effort, do you think you can say something about the effort?

Rule (f)

(16.23) Par: I don't know if he was bored or not, so I don't know if he .., I can't answer. Yeah, he answered them all, he attempted to answer all of them, well most of them, so he had an average effort, so I leave it there.

(16.47) Int: So, basically this is again, this is the type of feedback that the systems gives to him, so this one means the ones that are correct and this is the verbal feedback that is given to him so he only knows how many he got correct, how many he did wrong, and just "Well done, would you like to continue?" So now he has a choice to look at whether he did right or wrong and then he has a little bit of verbal feedback, so again, if you reckon that from this type of feedback if you think you could change any of these... you can comment about that.

(17.34) Par: I can’t change... Everything I said here about how he interacted, is what I, it is very difficult to change these around, because this is just a split second that the feedback came up here, you know, so, about the feedback, putting a tick at the end of the line, I don’t think it is a very good idea. The interface itself is quite difficult to take it all in, because the system mixes [?] numbers and there are not even in straight lines, personally I think it is about the interface. I don’t know it this is relevant to this but... No, I think, I can’t tell if his satisfaction level has changed, and

(18.32) Int: It is OK, I mean, if you cannot say anything, maybe something later, there are two more lessons. So, this just shows that the student is looking at it for a while,

(18.43) Par: Ah, so he is looking at it, OK

(18.46) Int: Because otherwise sometimes it looks like the system crashed.

(18.53) Int: So the student was looking a little bit to the feedback, then he clicked "yes" that he would like to continue and then this lesson is given, so is basically the same type of lesson, but now with the higher numbers, up to 100.

(19.10) Par: So, is he thinking now?

(19.13) Int: No, basically this is now to see if you can tell anything about these factors.

Rule (e)

(19.20) Par: Well, if he is continuing, no I still think that if he is a person that likes a lot of fantasy he is still bored. So I presume his sensory input is quite low.

(19.39) Int: Because the interface remains the same type?

(19.44) Par: Yes

(19.50) Par: He likes a lot of control and I think he has a lot of control, he is not time limited, so he can click

(20.01) Int: And also, well, the system offered him the possibility of whether he would like to continue or not, that is his choice.

(20.08) Par: I think he has relatively a high degree of control. Of course to have full control he could just see the written number and then maybe type in what he thought that number was, and not have a multiple choice, to have full control, but I think this is OK for him. I don't think it is very challenging for him, again because if he likes a lot of fantasy, this would encourage him, but I don't think this interface is very stimulating, so he might not find it very challenging. But then again, the task of learning numbers in a foreign language isn't very challenging anyway, it's hard to tell.

(20.51) Int: So, and then, given these you basically comment a little bit about these two, about sensory interest or cognitive interest. So sensory interest you think it would perhaps remain the same, very low, but cognitive interest is ...

(21.14) Par: I put cognitive interest up a little bit further than sensory interest because I think that he actually wants to learn.

(21.20) Int: Because he is continuing?

(21.21) Par: Because he is continuing, yes, and judging by the fact that he got all the other ones all-right, well, mostly all-right, so he is quite good, and he wants to continue, so in that respect, he is making a semi-effort into actually completing the task, so I think that is all I will say.

(21.20) Int: OK, so let's see how he does in this one. So this is actual time, OK, this is the time that he spent, because at the beginning I thought that the system crashed sometimes

(22.48) Int: So now, this is the same type of situation as before. I just ask you if you can update something there, just by default.

(22.56) Par: Do I know what type of help he is getting?

(22.58) Int: No, not yet. Later you will get. Is the same principle as before, you will get the help right after.

Rule (g)

(23.02) Par: OK, I think he is making a bigger effort at the moment, he is taking each line at a time, it seems, I only know this because he keeps the mouse there, whether he is or isn't I don't know, but it seems that he is taking one row at a time, and eliminating any of the incorrect answers.

(23.28) Int: How can you know that he is trying to eliminate the wrong answers?, is he moving?

(23.30) Par: Well, no, I presume that is what he is doing, maybe that is what I do, so it’s hard to tell, he could have his mouse up at the top, but he could be looking at something else, I don’t know that but it seems that, I think he moved the mouse across. And he also had the mouse in this ‘‘ju-ichi’’, so the ‘‘go’’ is the five, I think, so he was possibly trying to, I don’t know, the are all fifties, I don’t know, but I think he is making a bigger effort to get each one right.

(24.15) Int: Because it looks like he is going line per line?

Rule (h)

(24.16) Par: Yes, because he is going line by line and I think, in that respect, he is also more confident that he knows, that he doesn’t need to depend on the other possible answers for the other questions to help him answer the question at hand.

Rule (i)

(24.35) Int: And with this limited interface, how would you know that he is trying to depend on the other possible answers?

(24.40) Par: Well, like I said in the other time, he was moving around the mouse, looking at all the possible answers, say if he was trying to answer this top one he was looking, it seemed again, like he was looking around [?] trying to find some clue, yes, that’s the word, trying to find a clue that will help him answer, not the question where he got the clue from but the question he is trying to answer at that time, which could be the first one.

(25.14) Int: And from the fact that he is asking for some help?

Rule (j)

(25.16) Par: I think what happened, he was looking at 63 and 36, again, they are written quite similarly in Japanese, so he knows it is one or the other, but he is not too sure, but again he really wants to get it right, which tells me that he is a little bit more interested and also a little bit more cognitively interested.

(25.50) Int: Because it seems that he is doing an actual effort to do the exercise right?

(25.54) Par: Yes, yes, I think he wants to get it right. I'm leaving the sensory interest still low for the same reasons as before, and his satisfaction levels I don't know.

(26.05) Int: OK, so let's see the help that he gets .... sometimes is not very helpful the system actually.

(26.20) Par: [reading the help]

(26.40) Int: The help could tell you anything?

(26.45) Par: I think this is the type of help that he was looking for, because he had the mouse between 63 and 36 and if they tell you that ''ju'' is ten he should be OK, and it gives you an example as well, so by saying ''nana-ju'' here, which is 7 and a 10, and his possible answer will be 6 and a 10, and he, although, the one before he got 6 wrong, that might help him go back, so now he knows it is, unless he thinks that ''roku'' is three.

(27.25) Int: But at least he knows the principle now.

(27.29) Par: Yes, I think the feedback was good. I thought it was helpful, so. But I don't think that changes anything here.

(27.40) Int: OK, so let's see if he manages to make some use of the help.

(27.44) Par: He's given up.

(27.50) Int: So this means that he is now reading the feedback before he will press the ...

(27.54) Par: OK.

(28.16) Int: And at this point he gave up, so again if you can comment on the type of feedback that you would give, and whether you can infer something.

(28.28) Par: Maybe with the type of feedback. The first box he got I thought it was good, because I thought that was what he was looking for, yes, the help box, but maybe even with the help box, there could have been another box for more help, if he wanted to see the list of numbers from 60 to 70 or something to help him a bit more, almost like incrementally giving the answer away, but he is giving up now, so I don’t think he was very satisfied with the feedback he got, so I will move that down.

(29.15) Int: So at the beginning you said that it looked like helpful feedback, but the fact that he is giving up maybe makes you change your mind, that he is actually not satisfied with the whole thing?

Rules (k) and (l)

(29.30) Par: Yes, it’s hard to tell, but maybe he is still interested, I don’t know, I have to leave that one in the middle. I don’t know if his confidence was affected, he gave up, so must be lower. And he just gave up, so there was no effort. I’ll leave it here. I don’t know if he is going to go back again, you see, so.. So he is giving up completely, this is the end for him?

(30.12) Int: The end of this lesson.

(30.24) Par: But he just made no effort.

(30.26) Int: The system ticks the ones that were correct and says "Do you want to give up? You weren’t doing bad" which I think is not grammatically correct, but.

(30.40) Par: You were doing good, is perhaps more positive.

(30.49) Int: So, that is the feedback that he gets, and again if you think you can update something, otherwise we can see the next lesson.

Rule (m)

(31.00) Par: His sensory interest might go up a little bit, because the system is talking back to him, is speaking, is encouraging him, you know what I mean?

(31.20) Int: Not because the task itself, but because the system is responsive?

(31.24) Par: Some people find that really patronising, but I don't know, maybe he likes that.

(31.32) Int: But you consider maybe, it is something that the system is talking directly with him, is having a dialogue, perhaps is making him interested in the interface itself.

(31.35) Par: Exactly.

(31.48) Par: These are just the same, because I don't know.

(31.56) Int: So now again he is looking at the feedback at the answers he got, before he decides whether he would like to continue or not, which he says ''Yes''. So this is the third type of exercise that he gets, so this is just, he has to write the Japanese numbers in this boxes, and this is the solution for you in case you don't know, so he will have to type the number in there.

(32.33) Par: So, did he know that the task was going to be different?

(32.36) Int: No, he doesn't know what is coming next. The system follows some rules to decide which lessons to do next, but he doesn’t know in advance.

Rule (n)

(32.47) Par: Well, he has a lot of control here, I think, he likes to have control, so basically he has full control over what possible answers he can write, so maybe more satisfied with this type of task.

(33.00) Int: Because it reflects a bit more the type of control that he likes?

(33.07) Par: Yes, he has full control. As I said before, in the previous one it was multiple choice and he didn’t have full control, it was limited to four answers, so maybe with this one he has full control of what he is going to write in the box, whether it is right or wrong.

(33.30) Int: And the sensory interest, do you think it will change anything?

Rule (o)

(33.35) Par: I don’t like the interface myself, I don’t know, I will leave it there, just below average. I don’t know yet if he is cognitively interested or not. I leave his confidence as average, because he did go back, you know, although he was going to give up, so he is making an effort again there also.

(34.06) Int: So in the confidence, he gave up in the previous one, but then he said he would like to continue, so you think he is still confident, that he knows a little bit, and that the previous one was perhaps too difficult for him.

(34.15) Par: Yes, although this is a harder task, because he has to know exactly what to write, so his confidence, no, I leave it as average.

(34.30) Int: OK, I don't know if he does it properly here.

(35.40) Par: He is finished. Did he press "Done"?

(35.44) Int: He pressed "Done" just now.

(35.49) Par: Why he did that?, he was doing well. He mustn't be very satisfied with this, if he gave up half way through.

(36.08) Int: He has two options, to say he is finished or to say he gives up. Do you think that he choice of button that he chooses can tell you anything, whether he wants to give up or click "done"?

(36.30) Par: Well, give up should mean like I want to quit this task altogether, while "done" is more like this is as far as I can go, maybe the system can give him feedback to help him. The system knows that he only answered less than 50% of the questions, so maybe the system could know that he needs help with the others, but he didn't even try to answer the other ones?

(37.02) Int: So, how can you know that he didn't try?

(37.05) Par: Well, he had the mouse there, so he was obviously looking at this one, and he typed it in, and he followed the same pattern the whole way up, now if he knew 1 10 7 and then he got more difficult when he got to the higher numbers, but if he, I mean learning 1 to 10 he got those right, you know, he should know 5 and 9 I think, I presume that he didn't know 5 and 9.

(37.30) Int: So you are saying that it looks like he was doing and then when he got to the difficult ones he didn't even try those ones, because the mouse wasn't actually in this position, so it looks like he didn't really try.

(37.40) Par: Maybe, well, most people, if I was to answer this I would have my mouse there, I would indicate with my mouse where I was looking. But he, just after this one, he pressed done, don't know why. Maybe he did look at them, but it doesn't seem so, because the time that he finished writing in this one, and then very suddenly he just pressed "done". So it didn't seem like he gave too much thought.

(38.10) Int: So and that says anything about the others?

(38.31) Par: His effort was average, I think. I don't know. These are the same anyway. You see, I do that anyway, I indicate with my mouse where I am, so that's why I presume that when the mouse appears here he is thinking about this question or whatever.

(39.00) Int: OK, this is the last one. This is the same as before, the system says which ones are correct and the feedback is "OK, not too bad"

(39.15) Par: Can he continue with this actual exercise?

(39.19) Int: Well, he doesn't know yet. It says would you like to continue? The truth is that he will not be able to continue with this exercise, but he doesn't know yet.

(39.32) Par: Well, I'm presuming that he is going to give up but this exercise, because he only answered half so he is not making any effort, he is not making a huge effort

(39.42) Int: Because he gave up half way through?

(39.46) Par: Yes. I don't know about the ... I will have to see what he does next. If he says Yes or No I will be able to ...

(40.03) Int: So he says that he will like to continue, but the system presented a new lesson, it is the same type of lesson but with the higher numbers. But it is OK, if you want we can leave it here, already took a bit long.

(40.28) Par: Yes. Maybe he thought the other ones were too easy, and that's why he gave up. Ah he is giving up now. Ah he is thinking about it.

(40.43) Int: Looks like he is thinking about giving up.

(40.43) Par: Yes. So he is going to give up. He gave up. So I think his confidence has gone down, he know that these are too difficult.

(40.55) Int: Because he gave up.

(40.59) Par: He gave up. These are higher numbers and he didn't answer any, he didn't even try. He looked at them, but I think he thinks that these are too difficult for him, given that he didn't answer any of these and that he only answered half of the easier ones.

(41.20) Int: And that would reflect in any of the other factors?

(41.30) Par: I don’t think he is very satisfied. I will reduce that. I think maybe what he wanted was the next exercise, instead of being more difficult, just to be simpler numbers again. He should maybe have that option. The system should have said you want to continue with these numbers like numbers 20 down or do you want?

(41.55) Int: And why do you think he would be more satisfied if the system offered that option to continue?

(42.00) Par: Because then he can control the task, and he is a person that he likes a lot of control.

(42.15) Int: And here the system is forcing him.

(42.16) Par: Yes.

(42.17) Int: OK, let’s leave it here

B.3 Preliminary rules for motivation diagnosis study

Due to lack of space, it is not feasible to include in this appendix the complete transcription for all the interviews made as part of the Motivation Diagnosis study. In its place, we present here the first version of the Motivation Diagnosis Rules, which were extracted directly from the interviews. These rules represent a summary of the interviews, since we have attempted to present them using similar language and style to those used by the participants. A set of rules based on a more detailed analysis of these provisional rules is given in the main text of this dissertation in tables 5.4 to 5.13.

In order to make tables B.2 to B.11 easier to understand, we have followed the graphical conventions in table B.1[2] in order to represent the different parts that make each rule:

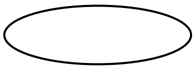
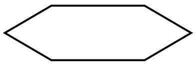
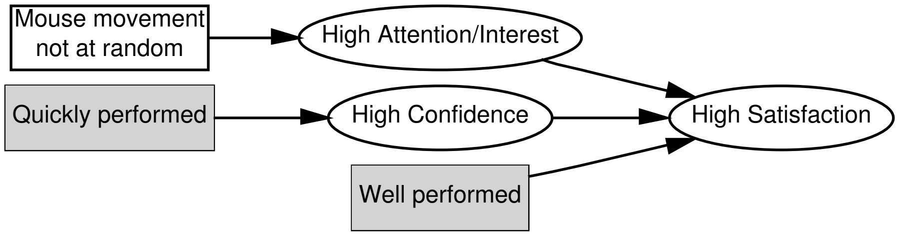
Node	Description
	Steps where the student's motivational models are involved
	Steps where interface issues are involved (e.g. moving the mouse a lot)
	Performance issues (e.g. time required to perform exercise)
	Other intermediate steps (e.g. student finds exercise harder)
	Steps involving feedback to be given to student
A $\rightarrow$ B	*A* implies *B*
A, B $\rightarrow$ C	*A* AND *B* imply *C*

Table B.1: Graphical conventions for tables B.2 to B.11

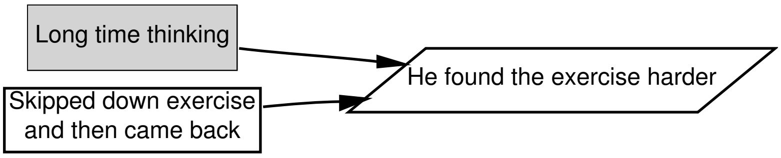
[2]This table appears in the main text of this dissertation as table 5.2

Table B.2: Motivation diagnosis rules of participant 1

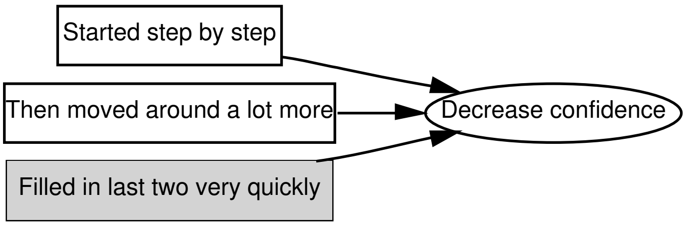
(a)



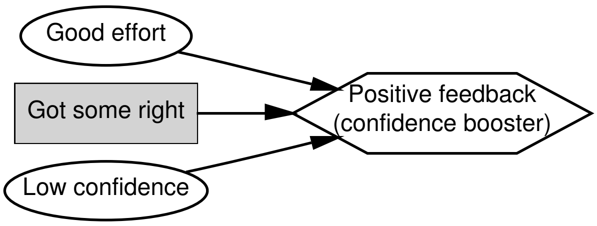
(b)



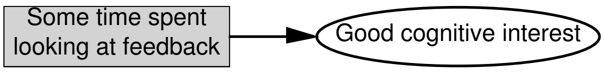
(c)



(d)



(e)



(f)

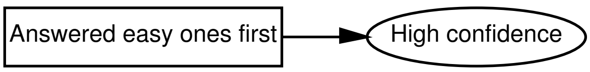
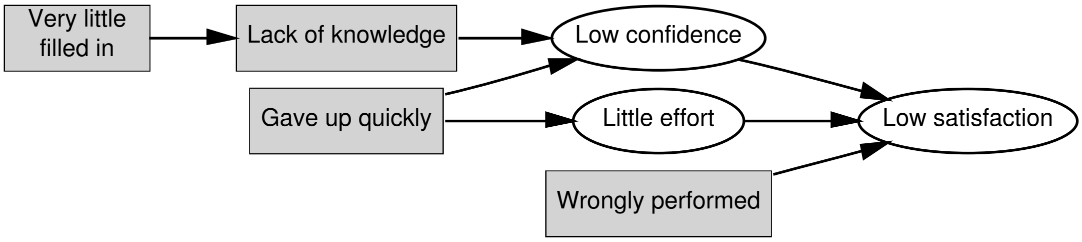
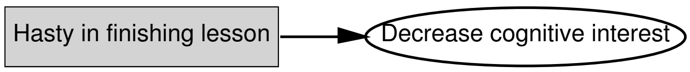


Table B.2: Motivation diagnosis rules of participant 1 (continued)

(g)



(h)



(i)

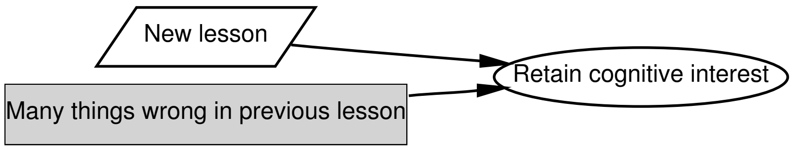
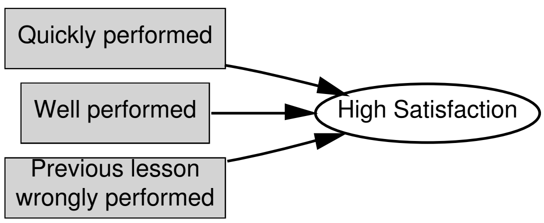


Table B.3: Motivation diagnosis rules of participant 2

(a)



(b)

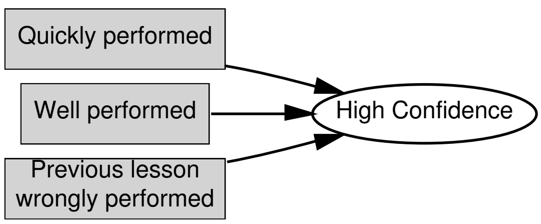
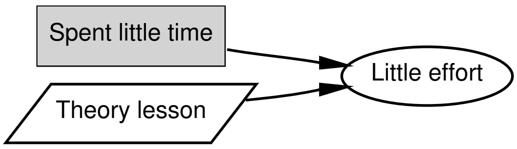
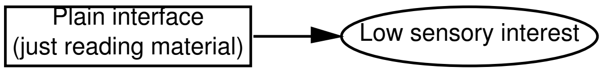


Table B.3: Motivation diagnosis rules of participant 2 (continued)

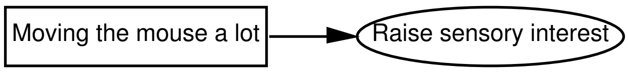
(c)



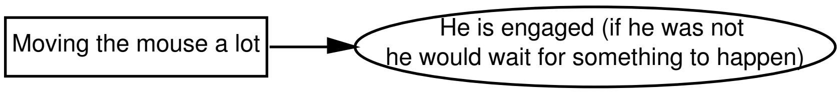
(d)



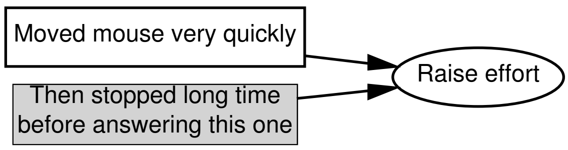
(e)



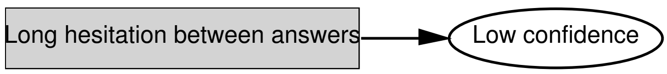
(f)



(g)



(h)



(i)

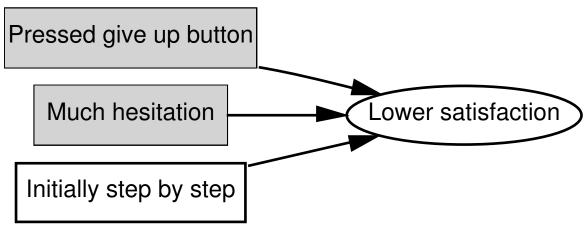
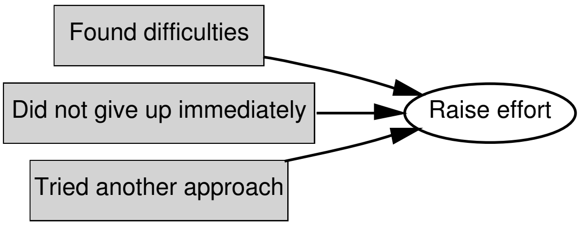
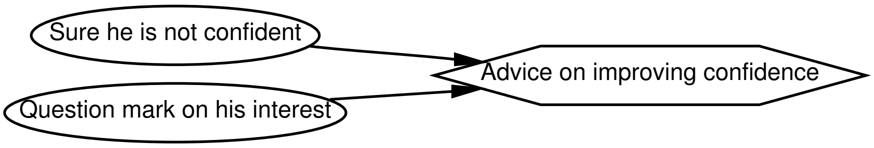


Table B.4: Motivation diagnosis rules of participant 3

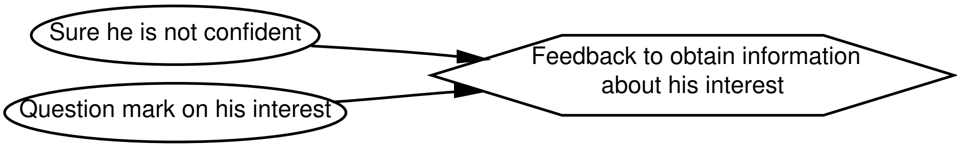
(a)



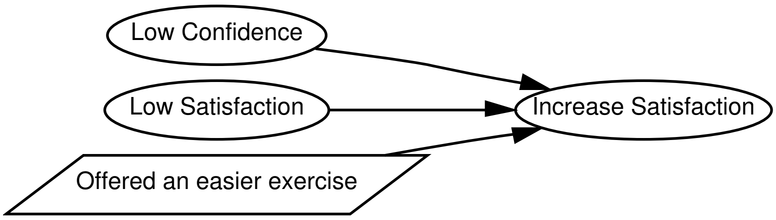
(b)



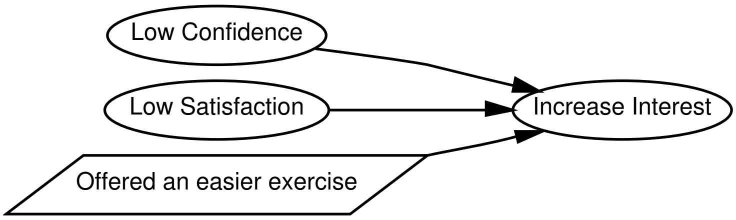
(c)



(d)



(e)



(f)

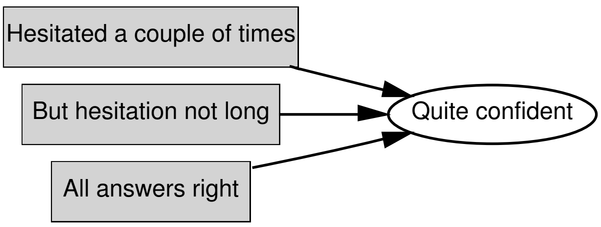
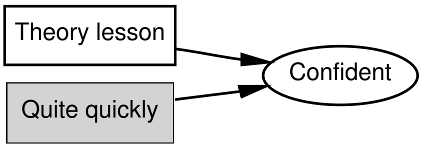


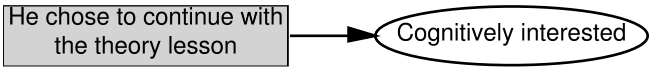
Table B.4: Motivation diagnosis rules of participant 3 (continued)

Table B.5: Motivation diagnosis rules of participant 4

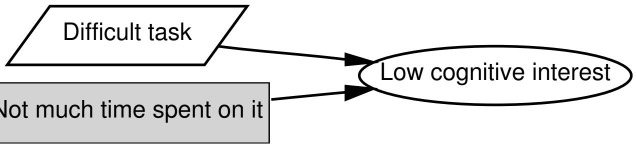
(a)



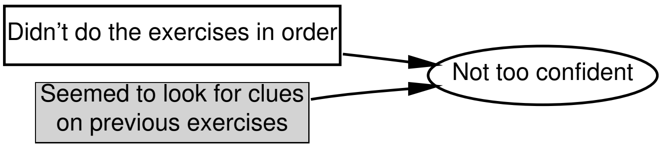
(b)



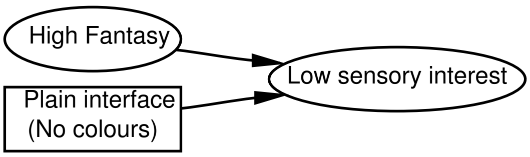
(c)



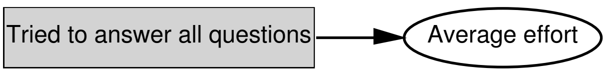
(d)



(e)



(f)



(g)

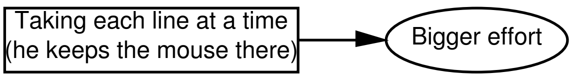
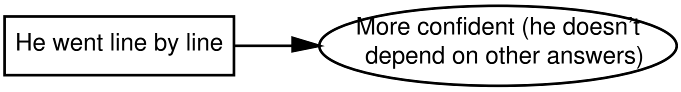
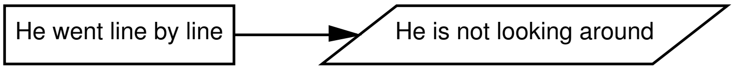


Table B.5: Motivation diagnosis rules of participant 4 (continued)

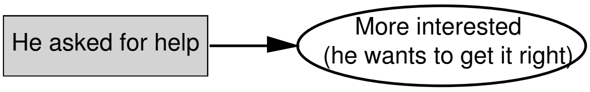
(h)



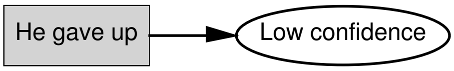
(i)



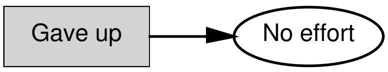
(j)



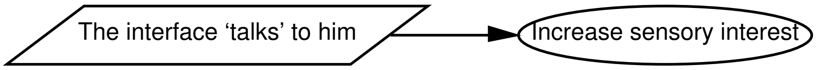
(k)



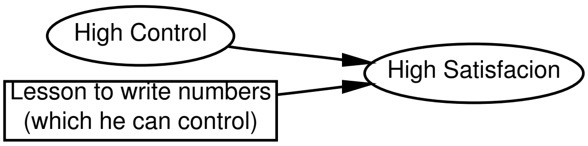
(l)



(m)



(n)



(o)

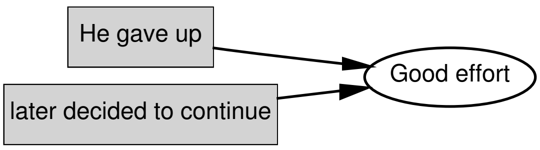
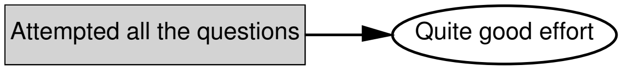
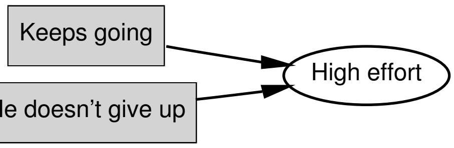


Table B.6: Motivation diagnosis rules of participant 5

(a)

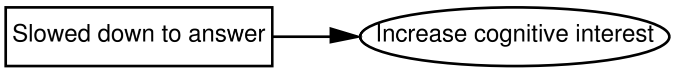


(b)

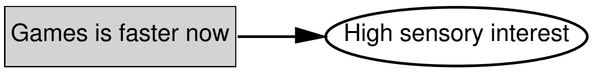


(c)

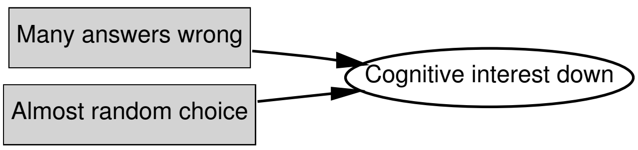
(d)



(e)



(f)



(g)

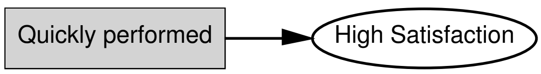


Table B.7: Motivation diagnosis rules of participant 6

(a)

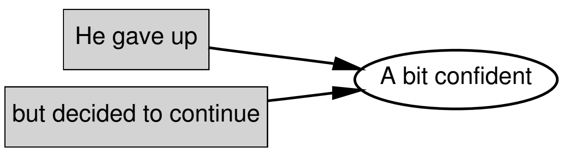
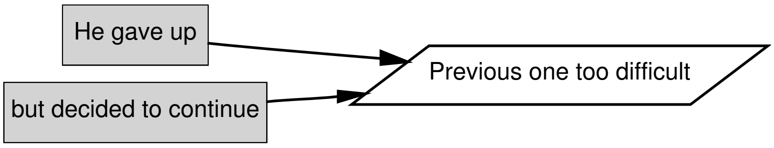
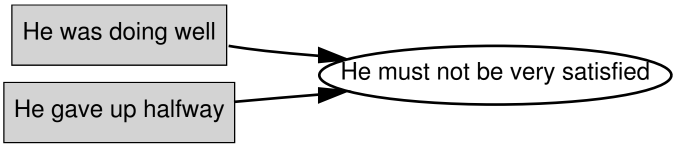


Table B.7: Motivation diagnosis rules of participant 6 (continued)

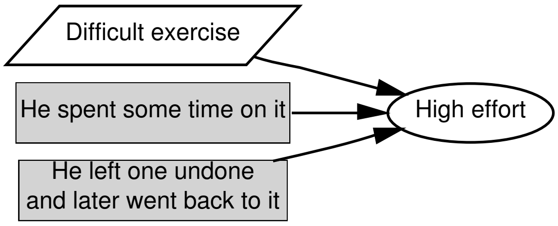
(b)



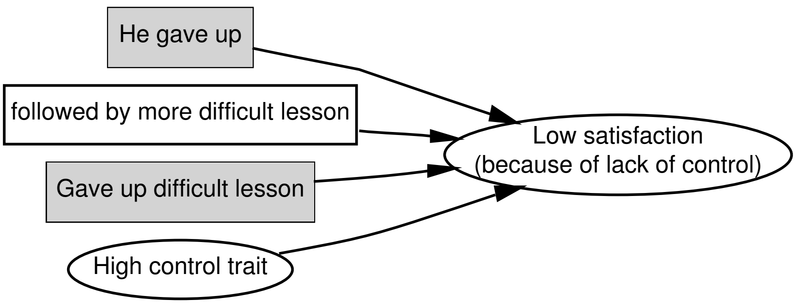
(c)



(d)



(e)



(f)

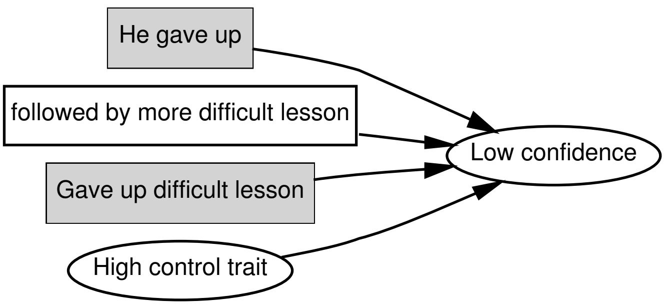


Table B.7: Motivation diagnosis rules of participant 6 (continued)

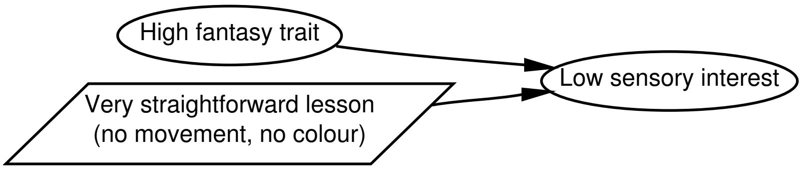
(g)	

Table B.8: Motivation diagnosis rules of participant 7

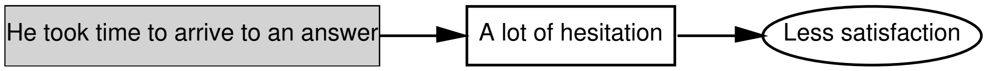
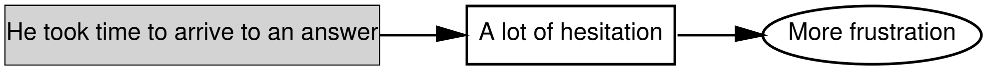
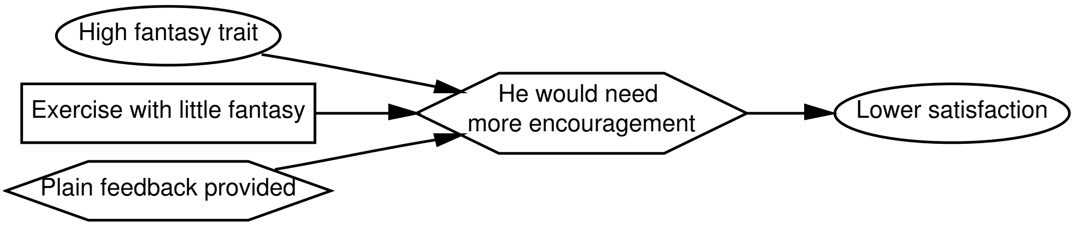
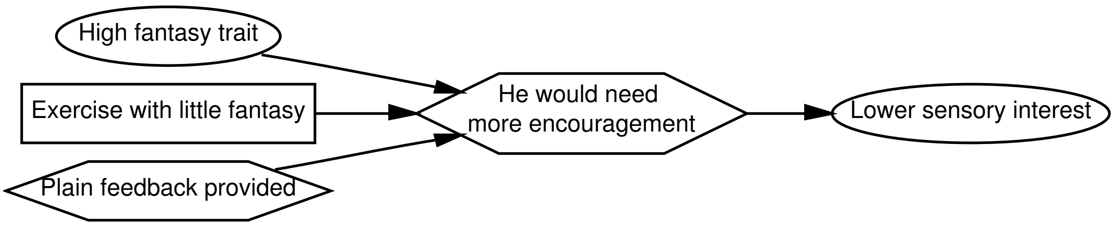
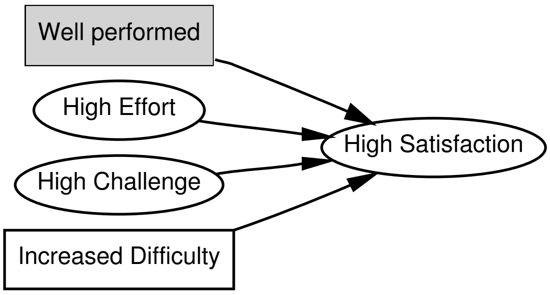
(a)	
(b)	
(c)	
(d)	

Table B.8: Motivation diagnosis rules of participant 7 (continued)

(e)



(f)

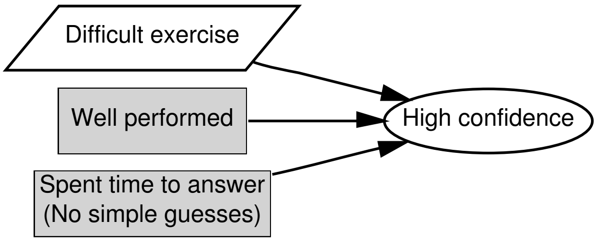
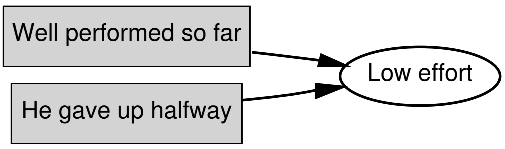
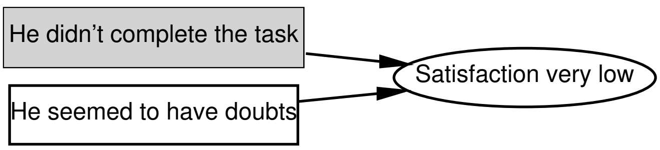


Table B.9: Motivation diagnosis rules of participant 8

(a)



(b)



(c)

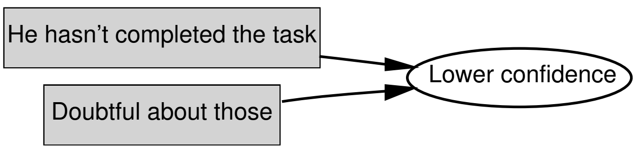
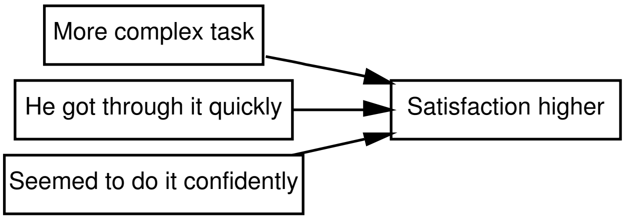
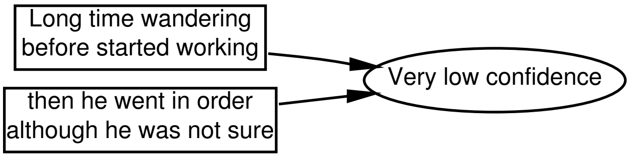


Table B.9: Motivation diagnosis rules of participant 8 (continued)

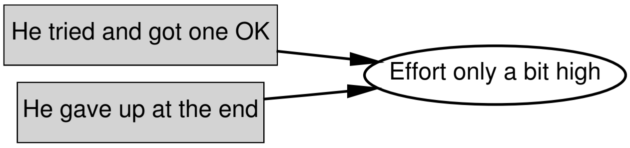
(d)



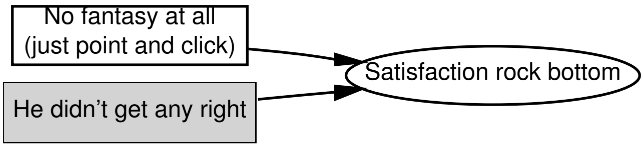
(e)



(f)



(g)



(h)

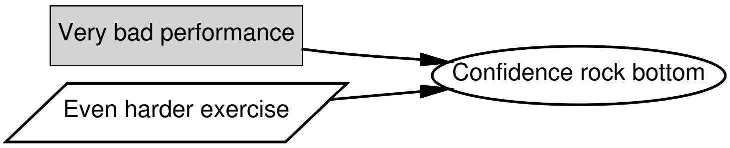
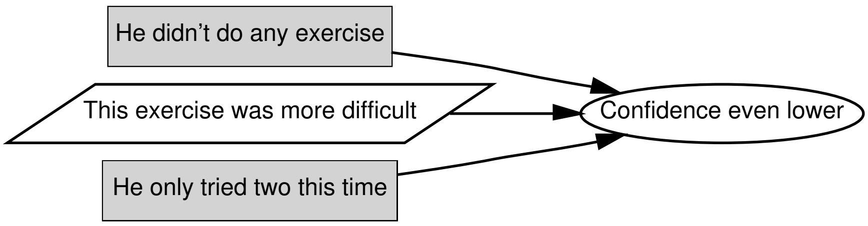
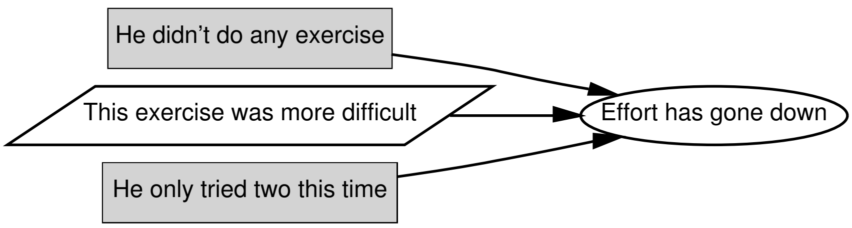


Table B.10: Motivation diagnosis rules of participant 9

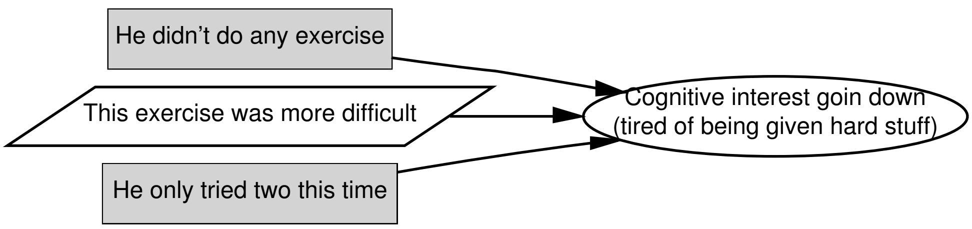
(a)



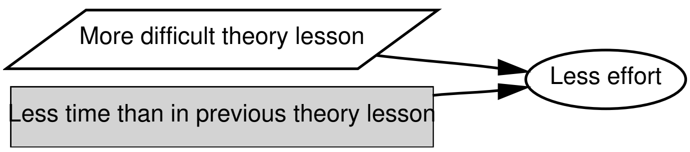
(b)



(c)



(d)



(e)

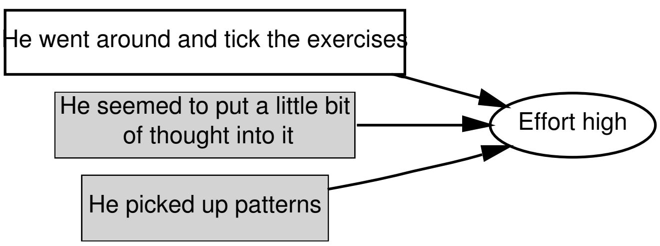


Table B.10: Motivation diagnosis rules of participant 9 (continued)

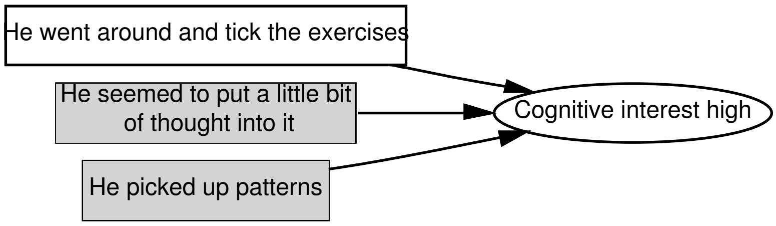
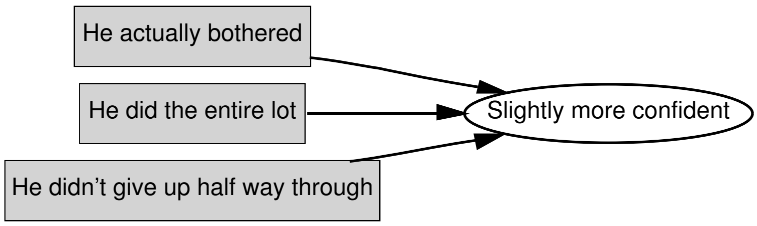
(f)	
(g)	

Table B.11: Motivation diagnosis rules of participant 10

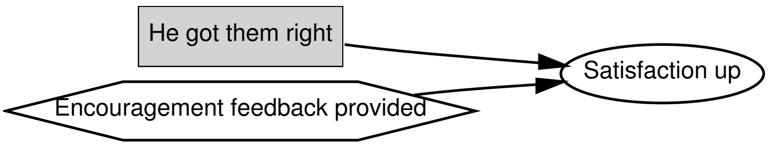
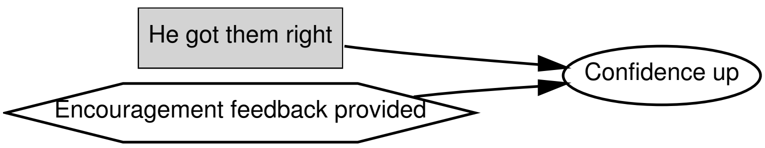
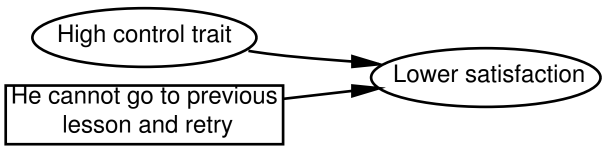
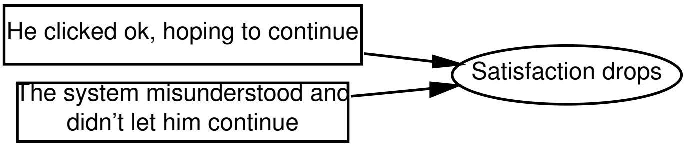
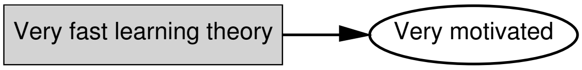
(a)	
(b)	
(c)	

Table B.11: Motivation diagnosis rules of participant 10 (continued)

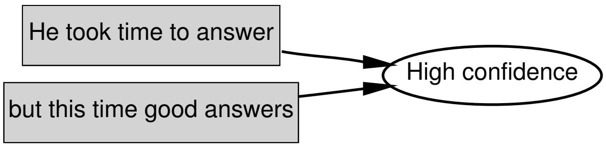
(d)



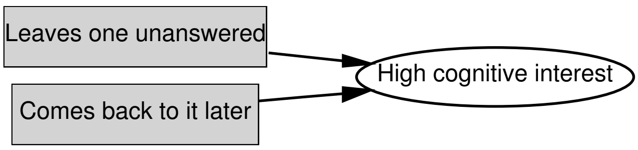
(e)



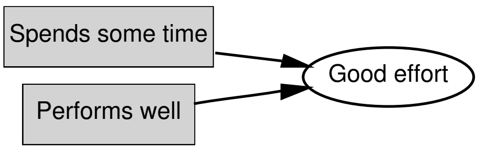
(f)



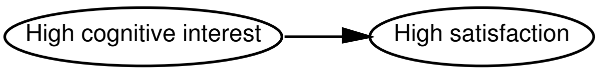
(g)



(h)



(i)



(j)

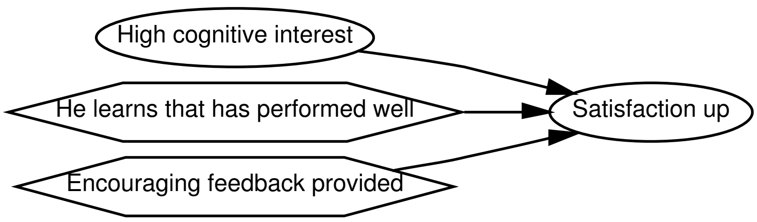
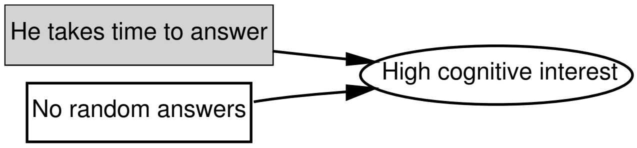


Table B.11: Motivation diagnosis rules of participant 10 (continued)

(k)



Appendix C

Validity study materials

This appendix contains the following materials in relation to the Validity study: questionnaire instructions, the complete questionnaire, and some of the results of the study in details.

C.1 Questionnaire instructions 272

C.2 Complete questionnaire 279

C.3 Validity study detailed results 288

C.1 Questionnaire instructions

Motivation Detection Study

Thanks for agreeing to participate in this study, which is divided in two main sections:

- Instructions
- Questionnaire

There are 4 pages of instructions (including this one), and it is important that you understand them all. Please read all of them carefully before proceeding to the questionnaire.

Overview

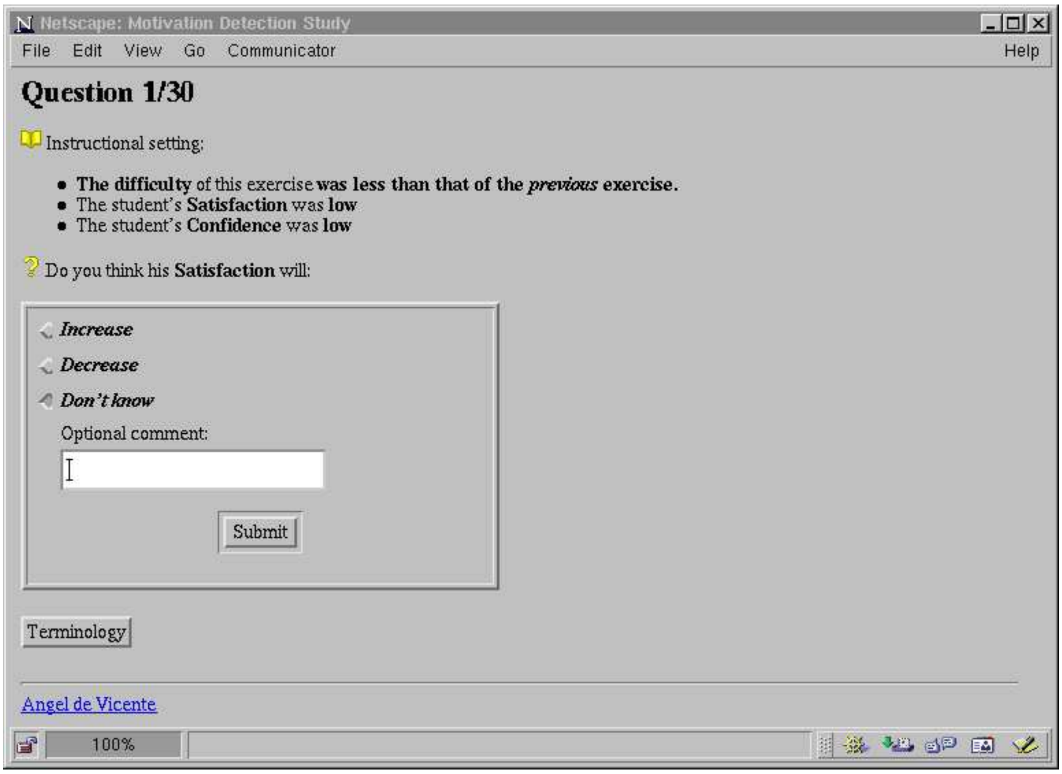
This study is part of on-going research on how computerised instructional systems can detect a student's motivational state (i.e. his/her motivation to study).

Your task in this study is to answer a number of questions about how a hypothetical instructional setting would affect a student's motivational state. Further details of your task are given in the following pages.

Angel de Vicente

Sample question

An example of the type of questions that you will have to answer is illustrated in the figure below.



Each question consists of 3 main parts:

1. Instructional Setting description.
2. Question.
3. Terminology button.

Instructional Setting description

The instructional setting description is given as a set of simple sentences that describe a number of different aspects of the instruction.

It is assumed that the instruction is a one-to-one computerised setting. That is, you should imagine a setting in which a student is using an instructional software, and the Instructional Setting description describes a variety of different aspects, such as the difficulty of the lesson presented, the time spent working on the lesson, etc. You should also imagine that the student has just finished performing a particular exercise, and it is at this point that you are asked to answer the question.

Question

The question always involves making a prediction about a particular motivational factor of the student given the presented instructional setting. Thus, in this example you are asked to predict whether you think the student's satisfaction would be high or low, given the presented instructional setting.

From the options given you should choose the one that you think is most appropriate, or you should choose "Don't know" if you cannot make a decision.

If you wish, you can also provide extra comments and/or qualifications of your answer in the text area provided.

Terminology button

Throughout the questionnaire we use a number of terms (which we introduce on the following page) whose definition should be clear to you. You need to read and understand all these terms, but it is not necessary for you to memorise them. On every question page you will see a button called "Terminology", which you can press at any time in order to see the definition of these terms.

Angel de Vicente

Terminology

Throughout the questionnaire, we use a number of terms. Some of them are not necessarily straightforward and we explain them on this page. You should try to understand all of them, but it is not important to memorise them, since you will be able to look at these definitions again at any time during the study. The terms are divided into three main categories:

1. Teaching materials
2. Motivational traits
3. Motivational state

Teaching materials

This category refers to characteristics of the lessons presented to the student, and you will encounter the following terms:

Term	Refers to:
Controllable	The degree of choices available to the student in a given lesson. (e.g. can the student select the order in which to do the tasks?)
Fantasy characteristics	The degree of imaginary items in the current lesson. (i.e. environments that evoke mental images of physical or social situations not actually present.)

Motivational traits

The motivational traits represent some stable characteristics of the student, which are not likely to change during the instructional interaction. These are:

Term	Refers to:
Control	The degree of control that the student likes having over the learning situation (i.e. does he like to select which exercises to do, in which order, etc. rather than let the instructor take these decisions?).
Challenge	The degree that the student enjoys having challenging situations during the instruction (i.e. does he like to try difficult exercises that represent a challenge for him?).
Fantasy	The degree that the student appreciates environments that evoke mental images of physical or social situations not actually present (i.e. does he like the learning materials being embedded in an imaginary context?).

Motivational state factors

These represent transient characteristics of the student (i.e. characteristics that are likely to change during the course of the instruction). These are the factors that you will be asked about in each question.

Term	**Refers to:**
Satisfaction	Refers to the overall feeling of goal accomplishment (i.e. does the student think that the instruction is satisfying and that it is getting him closer to his educational goals?).
Confidence	Refers to the student's belief in being able to perform the task at hand correctly.
Effort	Refers to the degree that the student is exerting himself in order to perform the learning activities.
Cognitive interest	Refers to curiosity aroused through the cognitive characteristics of the task (i.e. regardless of the presentation issues, does the student find the task at hand cognitively appealing?).
Sensory interest	Refers to the amount of curiosity aroused through the interface presentation (i.e. appeal of graphics, sounds, etc.).

Angel de Vicente

Summary

Let's summarise your task before proceeding with the questionnaire.

Once you start the questionnaire, you will be asked a number of questions, which will be presented in the context of an instructional setting.

The instructional setting consists of a number of descriptions about the task, student's performance, student's motivational traits, etc.

Given this instructional setting you will be asked to predict the likely value of a motivational factor by choosing from a number of options. If you are not sure you can always select the option "Don't know".

And remember you can always look at the terminology definitions by pressing the button "Terminology" available on every page.

That's all. Whenever you are ready, press the button "Questionnaire" to start.

Angel de Vicente

Please enter the following information

Name:	
E-mail (optional):	
Teaching experience (in years):	
Subjects taught:	
Username:	
Password:	

Angel de Vicente

C.2 Complete Questionnaire

Instructional setting: (**IS1**)

- The student performed the exercise well.
- The student's Confidence was high
- The student's Cognitive Interest was high

What do you think his Satisfaction level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**IS2**)

- The student performed the previous exercise badly.
- The student performed the exercise well.
- The student completed quickly the exercise.

What do you think his Satisfaction level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**IS3**)

- The student performed the exercise well.
- The student received encouragement feedback.

Do you think his Satisfaction will:

- Increase
- Decrease
- Don't know

Instructional setting: (**IS4**)

- The student performed the exercise well.
- The student received encouragement feedback.
- The student's Cognitive Interest was high

What do you think his Satisfaction level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**IS5**)

- The student's Cognitive Interest was high

What do you think his Satisfaction level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**IS6**)

- The student performed the exercise well.
- The difficulty of this exercise was greater than that of the previous exercise.
- The student's Effort has been high
- The student's Challenge trait value is high

What do you think his Satisfaction level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**IS7**)

- The student performed the exercise well.
- The student completed quickly the exercise.
- The difficulty of this exercise was greater than that of the previous exercise.

Do you think his Satisfaction will:

- Increase
- Decrease
- Don't know

Instructional setting: (**IS8**)

- The difficulty of this exercise was less than that of the previous exercise.
- The student's Satisfaction was low
- The student's Confidence was low

Do you think his Satisfaction will:

- Increase
- Decrease
- Don't know

Instructional setting: (**IS9**)

- The exercise was highly controllable.
- The student's Control trait value is high

What do you think his Satisfaction level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DS1**)

- The student performed the exercise badly.
- The student's Confidence was low
- The student's Effort has been low

What do you think his Satisfaction level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DS2**)

- The student's performance was satisfactory.
- The student gave up the exercise.

What do you think his Satisfaction level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DS3**)

- The student gave up the previous exercise.
- The student gave up the exercise.
- The difficulty of this exercise was greater than that of the previous exercise.
- The student's Control trait value is high

What do you think his Satisfaction level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DS4**)

- The student completed very slowly the exercise.

Do you think his Satisfaction will:

- Increase
- Decrease
- Don't know

Instructional setting: (**DS5**)

- The exercise offered small number of fantasy characteristics.
- The student received feedback.
- The student's Fantasy trait value is high

Do you think his Satisfaction will:

- Increase
- Decrease
- Don't know

Instructional setting: (**DS6**)

- The student performed the exercise very badly.
- The exercise offered very small number of fantasy characteristics.

What do you think his Satisfaction level will be at this point?

- Very High
- Very Low
- Don't know

Instructional setting: (**DS7**)

- The exercise was not controllable.
- The student's Control trait value is high

Do you think his Satisfaction will:

- Increase
- Decrease
- Don't know

Instructional setting: (**IC1**)

- The student performed the previous exercise badly.
- The student performed the exercise well.
- The student completed quickly the exercise.

What do you think his Confidence level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**IC2**)

- The student performed the exercise well.
- The student hesitated a little while doing the exercise.

What do you think his Confidence level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**IC3**)

- The student gave up the previous exercise.

What do you think his Confidence level will be at this point?

- High
- Average
- Don't know

Instructional setting: (**IC4**)

- The student performed the exercise well.
- The student completed the exercise on average time.
- The exercise was difficult.

What do you think his Confidence level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**IC5**)

- The student completed a large part of the exercise.
- The student did not give up the exercise.

Do you think his Confidence will:

- Increase
- Decrease
- Don't know

Instructional setting: (**IC6**)

- The student performed the exercise well.
- The student received encouragement feedback.

Do you think his Confidence will:

- Increase
- Decrease
- Don't know

Instructional setting: (**IC7**)

- The student performed the exercise well.
- The student completed the exercise on average time.

What do you think his Confidence level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DC1**)

- The student moved the mouse a lot during the exercise.
- The student hesitated a lot while doing the exercise.

Do you think his Confidence will:

- Increase
- Decrease
- Don't know

Instructional setting: (**DC2**)

- The student hesitated a lot while doing the exercise.

What do you think his Confidence level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DC3**)

- The student hesitated a lot while doing the exercise.
- The student did not perform the exercise in the order suggested.

What do you think his Confidence level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DC4**)

- The student gave up the exercise.

What do you think his Confidence level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DC5**)

- The student gave up the previous exercise.
- The student gave up the exercise.
- The difficulty of this exercise was greater than that of the previous exercise.
- The student's Control trait value is high

What do you think his Confidence level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DC6**)

- The student completed about half of the exercise.
- The student hesitated a lot while doing the exercise.

What do you think his Confidence level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DC7**)

- The student completed very slowly the exercise.
- The student hesitated a lot while doing the exercise.

What do you think his Confidence level will be at this point?

- Very High
- Very Low
- Don't know

Instructional setting: (**DC8**)

- The student performed the exercise very badly.
- The difficulty of this exercise was greater than that of the previous exercise.

What do you think his Confidence level will be at this point?

- Very High
- Very Low
- Don't know

Instructional setting: (**DC9**)

- The student completed a very small part of the exercise.
- The difficulty of this exercise was greater than that of the previous exercise.

Do you think his Confidence will:

- Increase
- Decrease
- Don't know

Instructional setting: (**IE1**)

- The student did not give up the exercise.
- The student hesitated a lot while doing the exercise.

Do you think his Effort will:

- Increase
- Decrease
- Don't know

Instructional setting: (**IE2**)

- The student completed a very large part of the exercise.

What do you think his Effort level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**IE3**)

- The student completed a large part of the exercise.
- The student did not give up the exercise.

What do you think his Effort level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**IE4**)

- The student performed the exercise in the order suggested.

Do you think his Effort will:

- Increase
- Decrease
- Don't know

Instructional setting: (**IE5**)

- The student gave up the previous exercise.

What do you think his Effort level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**IE6**)

- The student completed a large part of the exercise.
- The student completed the exercise on average time.
- The exercise was difficult.

What do you think his Effort level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**IE7**)

- The student performed the exercise well.
- The student completed slowly the exercise.

What do you think his Effort level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DE1**)

- The student completed quickly the exercise.
- The exercise was difficult.

What do you think his Effort level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DE2**)

- The student completed a large part of the exercise.

What do you think his Effort level will be at this point?

- High
- Average
- Don't know

Instructional setting: (**DE3**)

- The student gave up the exercise.

What do you think his Effort level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DE4**)

- The student performed the exercise well.
- The student completed about half of the exercise.
- The student gave up the exercise.

What do you think his Effort level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DE5**)

- The student's performance was satisfactory.
- The student completed a small part of the exercise.
- The student gave up the exercise.

What do you think his Effort level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DE6**)

- The student performed the exercise badly.
- The student completed a small part of the exercise.
- The difficulty of this exercise was greater than that of the previous exercise.

Do you think his Effort will:

- Increase
- Decrease
- Don't know

Instructional setting: (**DE7**)

- The student completed the exercise faster than the previous exercise
- The difficulty of this exercise was greater than that of the previous exercise.

Do you think his Effort will:

- Increase
- Decrease
- Don't know

Instructional setting: (**ICI1**)

- The student spent a long time looking at the feedback provided.

What do you think his Cognitive Interest level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**ICI2**)

- The difficulty of this exercise was less than that of the previous exercise.
- The student's Satisfaction was low
- The student's Confidence was low

Do you think his Cognitive Interest will:

- Increase
- Decrease
- Don't know

Instructional setting: (**ICI3**)

- The student asked for help during the exercise.

Do you think his Cognitive Interest will:

- Increase
- Decrease
- Don't know

Instructional setting: (**ICI4**)

- The student completed slowly the exercise.
- The student hesitated a little while doing the exercise.

What do you think his Cognitive Interest level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**ICI5**)

- The student performed the previous exercise badly.

What do you think his Cognitive Interest level will be at this point?

- High
- Average
- Don't know

Instructional setting: (**DCI1**)

- The student completed very quickly the exercise.

Do you think his Cognitive Interest will:

- Increase
- Decrease
- Don't know

Instructional setting: (**DCI2**)

- The student performed the exercise badly.
- The student hesitated a lot while doing the exercise.

Do you think his Cognitive Interest will:

- Increase
- Decrease
- Don't know

Instructional setting: (**DCI3**)

- The student completed quickly the exercise.
- The exercise was difficult.

What do you think his Cognitive Interest level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DCI4**)

- The student performed the exercise badly.
- The student completed a small part of the exercise.
- The difficulty of this exercise was greater than that of the previous exercise.

Do you think his Cognitive Interest will:

- Increase
- Decrease
- Don't know

Instructional setting: (**ISI1**)

- The student moved the mouse a lot during the exercise.

Do you think his Sensory Interest will:

- Increase
- Decrease
- Don't know

Instructional setting: (**ISI2**)

- The student asked to do again the exercise (which was presented as a game)

What do you think his Sensory Interest level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**ISI3**)

- This and the previous exercises were presented as a game.
- The game was played faster this time.

What do you think his Sensory Interest level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**ISI4**)

- The student received feedback.

Do you think his Sensory Interest will:

- Increase
- Decrease
- Don't know

Instructional setting: (**DSI1**)

- The exercise had no graphics, colours or sounds.

What do you think his Sensory Interest level will be at this point?

- High
- Low
- Don't know

Instructional setting: (**DSI2**)

- The student's Fantasy trait value is high
- The exercise had very little graphics, colours or sounds.

What do you think his Sensory Interest level will be at this point?

- High
- Low
- Don't know

C.3 Validity study detailed results

We can see a complete list of all the results of the questionnaire in table C.1. The columns represent respectively: the name of the rule (see tables 5.4 to 5.13); the output value of the rules; the number of participants who answered the question corresponding to that rule; the value of the first choice given to participants; the number of participants that chose that choice; the same for the second and third options.

Table C.1: Results of motivation diagnosis questionnaire.

Rule	Output	n	Option 1		Option 2		Option 3	
IS1	High	18	High	16	Low	0	Don't know	2
IS2	High	15	High	12	Low	0	Don't know	3
IS3	Increase	18	Increase	16	Decrease	0	Don't know	2
IS4	High	15	High	14	Low	0	Don't know	1
IS5	High	15	High	8	Low	1	Don't know	6
IS6	High	17	High	16	Low	0	Don't know	1
IS7	Increase	16	Increase	14	Decrease	1	Don't know	1
IS8	Increase	16	Increase	7	Decrease	6	Don't know	3
IS9	High	18	High	13	Low	1	Don't know	4
DS1	Low	15	High	0	Low	15	Don't know	0
DS2	Low	16	High	0	Low	12	Don't know	4
DS3	Low	17	High	0	Low	16	Don't know	1
DS4	Decrease	16	Increase	3	Decrease	2	Don't know	11
DS5	Decrease	18	Increase	8	Decrease	4	Don't know	6
DS6	Very low	15	Very High	0	Very Low	12	Don't know	3
DS7	Decrease	17	Increase	3	Decrease	10	Don't know	4
IC1	High	17	High	13	Low	0	Don't know	4
IC2	High	15	High	12	Low	0	Don't know	3
IC3	Average	18	High	1	Average	10	Don't know	7

continued on next page

continued from previous page

Rule	Output	n	Option 1		Option 2		Option 3	
IC4	High	15	High	13	Low	1	Don’t know	1
IC5	Increase	17	Increase	12	Decrease	0	Don’t know	5
IC6	Increase	15	Increase	14	Decrease	0	Don’t know	1
IC7	High	17	High	13	Low	0	Don’t know	4
DC1	Decrease	18	Increase	0	Decrease	9	Don’t know	9
DC2	Low	15	High	1	Low	10	Don’t know	4
DC3	Low	16	High	1	Low	13	Don’t know	2
DC4	Low	15	High	0	Low	12	Don’t know	3
DC5	Low	15	High	1	Low	12	Don’t know	2
DC6	Low	15	High	0	Low	12	Don’t know	3
DC7	Very low	15	Very High	0	Very Low	8	Don’t know	7
DC8	Very low	16	Very High	0	Very Low	15	Don’t know	1
DC9	Decrease	16	Increase	0	Decrease	14	Don’t know	2
IE1	Increase	15	Increase	12	Decrease	1	Don’t know	2
IE2	High	16	High	11	Low	0	Don’t know	5
IE3	High	16	High	11	Low	0	Don’t know	5
IE4	Increase	15	Increase	2	Decrease	1	Don’t know	12
IE5	High	15	High	0	Low	10	Don’t know	5
IE6	High	16	High	12	Low	1	Don’t know	3
IE7	High	16	High	14	Low	0	Don’t know	2
DE1	Low	16	High	6	Low	6	Don’t know	4
DE2	Average	17	High	8	Average	0	Don’t know	9
DE3	Low	15	High	0	Low	14	Don’t know	1
DE4	Low	15	High	0	Low	12	Don’t know	3
DE5	Low	18	High	0	Low	15	Don’t know	3
DE6	Decrease	15	Increase	1	Decrease	12	Don’t know	2
DE7	Decrease	15	Increase	10	Decrease	3	Don’t know	2

continued on next page

continued from previous page

Rule	Output	n	Option 1		Option 2		Option 3	
ICI1	High	17	High	15	Low	0	Don't know	2
ICI2	Increase	18	Increase	2	Decrease	12	Don't know	4
ICI3	Increase	15	Increase	12	Decrease	1	Don't know	2
ICI4	High	17	High	5	Low	6	Don't know	6
ICI5	Average	18	High	1	Average	11	Don't know	6
DCI1	Decrease	15	Increase	4	Decrease	5	Don't know	6
DCI2	Decrease	15	Increase	2	Decrease	11	Don't know	2
DCI3	Low	15	High	7	Low	2	Don't know	6
DCI4	Decrease	16	Increase	2	Decrease	10	Don't know	4
ISI1	Increase	15	Increase	7	Decrease	4	Don't know	4
ISI2	High	15	High	10	Low	0	Don't know	5
ISI3	High	15	High	13	Low	0	Don't know	2
ISI4	Increase	15	Increase	12	Decrease	0	Don't know	3
DSI1	Low	16	High	0	Low	14	Don't know	2
DSI2	Low	17	High	3	Low	12	Don't know	2

In table C.2 we can see the complete set of results of applying the chi-square 'goodness of fit' test to each rule. The columns represent respectively: the name of the rule; the number of participants who answered the corresponding question; the number of *accept* replies; the number of *reject* replies; the probability of obtaining that distribution of replies assuming the null hypothesis.

Table C.2: Chi-square results for all the rules.

Rule	n	Accept	Reject	p
IS1	18	16	2	<0.00001
IS2	15	12	3	0.00013
IS3	18	16	2	<0.00001

continued on next page

continued from previous page

Rule	n	Accept	Reject	p
IS4	15	14	1	<0.00001
IS5	15	8	7	0.10035
IS6	17	16	1	<0.00001
IS7	16	14	2	<0.00001
IS8	16	7	9	0.37387
IS9	18	13	5	0.00047
DS1	15	15	0	<0.00001
DS2	16	12	4	0.00038
DS3	17	16	1	<0.00001
DS4	16	2	14	0.07555
DS5	18	4	14	0.31731
DS6	15	12	3	0.00013
DS7	17	10	7	0.02502
IC1	17	13	4	0.00015
IC2	15	12	3	0.00013
IC3	18	10	8	0.04550
IC4	15	13	2	0.00001
IC5	17	12	5	0.00106
IC6	15	14	1	<0.00001
IC7	17	13	4	0.00015
DC1	18	9	9	0.13361
DC2	15	10	5	0.00617
DC3	16	13	3	0.00004
DC4	15	12	3	0.00013
DC5	15	12	3	0.00013
DC6	15	12	3	0.00013
DC7	15	8	7	0.10035

continued on next page

continued from previous page

Rule	n	Accept	Reject	p
DC8	16	15	1	<0.00001
DC9	16	14	2	<0.00001
IE1	15	12	3	0.00013
IE2	16	11	5	0.00252
IE3	16	11	5	0.00252
IE4	15	2	13	0.10035
IE5	15	0	15	0.00617
IE6	16	12	4	0.00038
IE7	16	14	2	<0.00001
DE1	16	6	10	0.72064
DE2	17	0	17	0.00338
DE3	15	14	1	<0.00001
DE4	15	12	3	0.00013
DE5	18	15	3	<0.00001
DE6	15	12	3	0.00013
DE7	15	3	12	0.27333
ICI1	17	15	2	<0.00001
ICI2	18	2	16	0.04550
ICI3	15	12	3	0.00013
ICI4	17	5	12	0.72844
ICI5	18	11	7	0.01242
DCI1	15	5	10	1.00000
DCI2	15	11	4	0.00102
DCI3	15	2	13	0.10035
DCI4	16	10	6	0.01286
ISI1	15	7	8	0.27332
ISI2	15	10	5	0.00617

continued on next page

continued from previous page				
Rule	n	Accept	Reject	p
ISI3	15	13	2	0.00001
ISI4	15	12	3	0.00013
DSI1	16	14	2	<0.00001
DSI2	17	12	5	0.00106

In table C.3 we can see the rules for which we cannot reject the null hypothesis (p>=0.01), and thus these are the rules that are not accepted for our MOODS prototype.

Rule	n	Accept	Reject	p
IS5	15	8	7	0.10035
IS8	16	7	9	0.37387
DS4	16	2	14	0.07555
DS5	18	4	14	0.31731
DS7	17	10	7	0.02502
IC3	18	10	8	0.04550
DC1	18	9	9	0.13361
DC7	15	8	7	0.10035
IE4	15	2	13	0.10035
DE1	16	6	10	0.72064
DE7	15	3	12	0.27333
ICI2	18	2	16	0.04550
ICI4	17	5	12	0.72844
ICI5	18	11	7	0.01242
DCI1	15	5	10	1.00000
DCI3	15	2	13	0.10035
DCI4	16	10	6	0.01286
ISI1	15	7	8	0.27332

Table C.3: Rules on which null hypothesis cannot be rejected.

The rules in table C.4 are exceptional, as the null hypothesis can be rejected ($p<0.01$), but they cannot be accepted. This is so since all the participants rejected the rule.

Rule	n	Accept	Reject	p
IE5	15	0	15	0.00617
DE2	17	0	17	0.00338

Table C.4: Null hypothesis can be rejected, but rules cannot be accepted.

Appendix D

Evaluation materials

This appendix contains the following materials in relation to the Evaluation study: questionnaire instructions, the complete questionnaire, and the detailed results for both the motivational model questions and the importance questions of the questionnaire.

D.1 Questionnaire instructions 296

D.2 Complete questionnaire 302

D.3 Motivational model question results 306

D.4 Importance question results 310

D.1 Questionnaire instructions

Motivation in Education Study

Thanks for agreeing to participate in this study, which is divided in two main sections:

- Instructions
- Questionnaire

There are 5 pages of instructions (including this one), and it is important that you understand them all. The instructions are a little bit long, but please read all of them carefully before proceeding to the questionnaire.

Overview

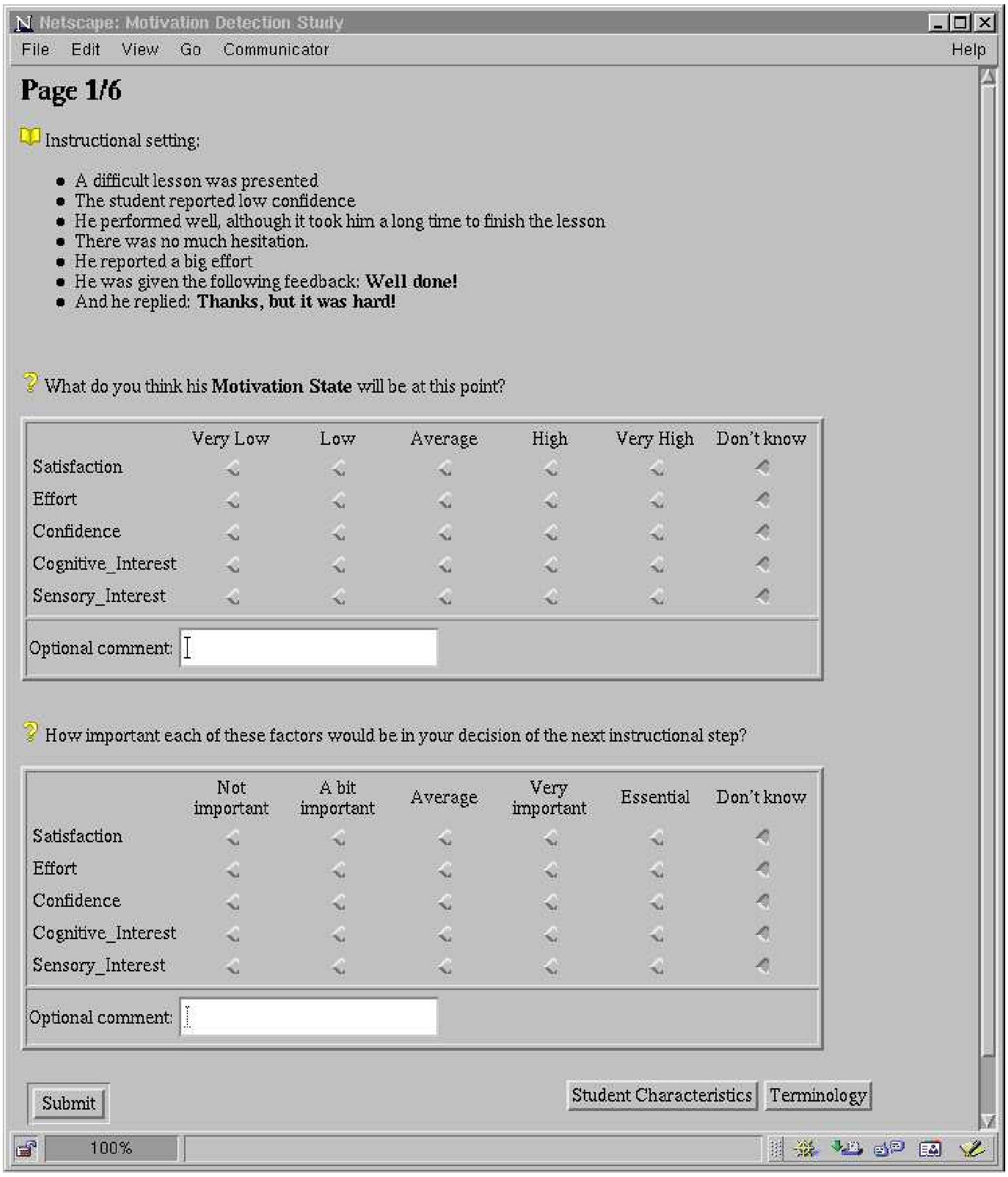
This study is part of on-going research on how computerised instructional systems can detect a student's motivational state (i.e. his/her motivation to study).

Your task in this study is to answer a number of questions about how a hypothetical instructional setting would affect a student's motivational state. Further details of your task are given in the following pages.

Angel de Vicente

Sample question

An example of the type of questions that you will have to answer is illustrated in the figure below.



Each question consists of 4 parts:

1. Instructional Setting description.
2. Question about student's Motivational state.
3. Question about the importance of Motvational Factors.
4. Row of buttons.

Instructional Setting description

The instructional setting description is given as a set of simple sentences that describe a number of different aspects of the instruction. Throughout the study you will be presented with a number of Instructional Settings, and you should imagine that these are consecutive in time. That is, you should imagine that the different settings that you will be presented with are sequential and that they constitute an interaction of a student with a learning system.

It is assumed that the instruction is a one-to-one computerised setting. That is, you should imagine a setting in which a student is using an instructional software, and the Instructional Setting description describes a variety of different aspects, such as the difficulty of the lesson presented, the time spent working on the lesson, student's performance, the feedback provided, etc.

Question about student's Motivational state.

The first question always involves making a prediction about the motivational state of the student given the presented instructional setting. For each of the given factors you should choose the value that you think is most appropriate, or you should choose "Don't know" if you cannot make a decision about that particular factor.

If you wish, you can also provide extra comments and/or qualifications of your answer in the text area provided.

Question about the importance of Motivational Factors.

The second question refers to the importance that you would give to each of the motivational factors if you had to decide which step to take next.

That is, given the presented instructional setting and given your prediction for the student's Motivational State, you should imagine that you are in control of the instruction. In order to decide what step to take next, which of the motivational factors would you consider essential, not important, etc?

If you wish, you can also provide extra comments and/or qualifications of your answer in the text area provided.

Row of buttons

Throughout the questionnaire we use a number of terms (which we introduce on the following page) whose definition should be clear to you. You need to read and understand all these terms, but it is not necessary for you to memorise them. On every question page you will see a button called "Terminology", which you can press at any time in order to see the definition of these terms.

Similarly, the button "Student Characteristics" will allow you to see the hypothetical student's characteristics at any time during the study.

Terminology

Throughout the questionnaire, we use a number of terms. Some of them are not necessarily straightforward and we explain them on this page. You should try to understand all of them, but it is not important to memorise them, since you will be able to look at these definitions again at any time during the study. The terms are divided into three main categories:

1. Teaching materials
2. Student characteristics
3. Motivational state

Teaching materials

This category refers to characteristics of the lessons presented to the student, and you will encounter the following terms:

Term	**Refers to:**
Controllable	The degree of choices available to the student in a given lesson. (e.g. can the student select the order in which to do the tasks?)
Fantasy characteristics	The degree of imaginary items in the current lesson. (i.e. environments that evoke mental images of physical or social situations not actually present.)

Student characteristics

These represent some stable characteristics of the student, which are not likely to change during the instructional interaction. They are:

Term	**Refers to:**
Control	The degree of control that the student likes having over the learning situation (i.e. does he like to select which exercises to do, in which order, etc. rather than let the instructor take these decisions?).
Challenge	The degree that the student enjoys having challenging situations during the instruction (i.e. does he like to try difficult exercises that represent a challenge for him?).
Fantasy	The degree that the student appreciates environments that evoke mental images of physical or social situations not actually present (i.e. does he like the learning materials being embedded in an imaginary context?).

Motivational state factors

These represent transient characteristics of the student (i.e. characteristics that are likely to change during the course of the instruction). These are the factors that you will be asked about in each question.

Term	**Refers to:**
Satisfaction	Refers to the overall feeling of goal accomplishment (i.e. does the student think that the instruction is satisfying and that it is getting him closer to his educational goals?).
Confidence	Refers to the student's belief in being able to perform the task at hand correctly.
Effort	Refers to the degree that the student is exerting himself in order to perform the learning activities.
Cognitive interest	Refers to curiosity aroused through the cognitive characteristics of the task (i.e. regardless of the presentation issues, does the student find the task at hand cognitively appealing?).
Sensory interest	Refers to the amount of curiosity aroused through the interface presentation (i.e. appeal of graphics, sounds, etc.).

Angel de Vicente

Summary

Let's summarise your task before proceeding with the questionnaire.

First, you will be presented with the characteristics of a hypothetical student. You should imagine that you are dealing with this student throughout the study.

Once you start the questionnaire, you will be asked a number of questions, which will be presented in the context of an instructional setting. The instructional setting consists of a number of descriptions about the task, student's performance, student's characteristics, etc.

Given this instructional setting you will be asked to predict the likely value of some motivational factors by choosing from a number of options. If you are not sure you can always select the option "Don't know". You will also be asked to comment about the importance that you would give to of each of these factors if you had to take the next instructional step.

You should imagine that the study is sequential. That is, that every question is consecutive in time to the previous one. And remember that you can always look at the terminology definitions by pressing the button "Terminology" available on every page. Similarly, you can also look at the Student Characteristics by pressing the corresponding button, which is also available on every page.

That's all. Whenever you are ready, press the button "Questionnaire" to start.

Angel de Vicente

D.2 Complete questionnaire

D.2.1 Simulation 1.

1

He performed the lesson very well.

He completed most of the exercises in the lesson.

He showed little hesitation.

He didn't ask for help.

He reported that he put a lot of effort in doing the exercises.

He was given the feedback: That was excellent! You did very well.

2

This time he was given a more difficult lesson.

He reported he was very confident that he could do this lesson well.

He performed the lesson well.

He completed all the exercises in the lesson.

He didn't complete the exercises in the given order.

He didn't give up the lesson.

He was given the feedback: That was very good.

3

We know that he is a person that enjoys a high degree of control.

This time he was given a more difficult lesson.

He completed about half of the exercises in the lesson.

Of those exercises that he completed, he performed the lesson satisfactorily.

He submitted the lesson quickly.

He didn't complete the exercises in the given order.

He was given the feedback: That was OK.

4

This time he was given a lesson of similar difficulty.

He submitted the lesson very quickly.

He showed little hesitation.

He performed the lesson poorly.

He didn't give up the lesson.

He was given the feedback: Sorry, you did poorly.

He spent some time looking at the corrections to his answers.

5

This time he was given a lesson of similar difficulty.

He reported that his cognitive interest was low.

He submitted the lesson quickly.

He performed the lesson poorly.

He reported that he put very little effort in doing the exercises.

He was given the feedback: Sorry, you did poorly.

He spent little time looking at the corrections to his answers.

6

We know that he is a person that enjoys a high degree of fantasy.

This time he was given a more difficult lesson.

The lesson had no fantasy characteristics.

He reported that he was confident that he could do this lesson.

He completed very few of the exercises in the lesson.

He decided to give up the lesson very quickly.

D.2.2 Simulation 2.

1

He performed the lesson very well.

He completed all the exercises in the lesson.

He submitted the lesson quickly.

He showed little hesitation.

He didn't give up the lesson.

He reported that he put average effort in doing the exercises.

$\boxed{2}$

We know that he is a person that enjoys high challenge.

This time he was given a more difficult lesson.

He performed the lesson well.

He completed most of the exercises in the lesson.

He completed the exercises in the given order.

He was given the feedback: That was OK.

$\boxed{3}$

This time he was given a more difficult lesson.

He reported he was very confident that he could do this lesson well.

He completed about half of the exercises in the lesson.

Of those exercises that he completed, he performed the lesson well.

He submitted the lesson quickly.

He reported that he put little effort in doing the exercises.

He was given the feedback: That was OK. Let's try something much harder now!

$\boxed{4}$

This time he was given a much more difficult lesson.

He reported that his cognitive interest was high.

He reported he was more or less confident that he could do this lesson well.

He performed the lesson well.

He submitted the lesson in average time.

He showed little hesitation.

He completed the exercises in the given order.

$\boxed{5}$

This time he was given a more difficult lesson.

He performed the lesson satisfactorily.

He completed all the exercises in the lesson.

He showed little hesitation.

He didn't give up the lesson.

He was given the feedback: That was OK.

6

```
This time he was given a more difficult lesson.
The given lesson was highly controllable.
He performed the lesson well.
He completed all the exercises in the lesson.
He didn’t give up the lesson.
He reported that he put a lot of effort in doing the exercises.
He was given the feedback: That was very good. You did very well.
```

D.3 Motivational model question results

We can see a complete list of all the results for the question regarding the values of the motivational model in the questionnaire (first question) in tables D.1 and D.2.

Table D.1: Results of motivational model question, Simulation 1.

SITUATION 1

	Very low	Low	Average	High	Very High	Don't know
Satisfaction	0	0	1	4	14	0
Effort	0	1	2	8	7	1
Confidence	0	0	0	5	14	0
Cognitive Interest	0	0	1	6	8	4
Sensory Interest	0	0	3	5	5	6

SITUATION 2

	Very low	Low	Average	High	Very High	Don't know
Satisfaction	0	2	2	8	7	0
Effort	1	1	1	7	8	1
Confidence	0	1	5	8	5	0
Cognitive Interest	1	0	2	7	7	2
Sensory Interest	1	0	6	3	6	3

SITUATION 3

	Very low	Low	Average	High	Very High	Don't know
Satisfaction	1	5	12	1	0	0
Effort	0	6	8	4	0	1
Confidence	0	6	10	2	0	1
Cognitive Interest	0	6	6	4	0	3
Sensory Interest	1	5	5	2	0	6

continued on next page

continued from previous page

SITUATION 4

	Very low	Low	Average	High	Very High	Don't know
Satisfaction	4	13	1	1	0	0
Effort	2	8	6	1	1	1
Confidence	2	13	2	1	0	1
Cognitive Interest	1	8	4	2	2	2
Sensory Interest	2	7	4	1	1	4

SITUATION 5

	Very low	Low	Average	High	Very High	Don't know
Satisfaction	8	11	0	0	0	0
Effort	8	11	0	0	0	0
Confidence	7	9	1	0	1	1
Cognitive Interest	9	10	0	0	0	0
Sensory Interest	6	9	0	0	0	4

SITUATION 6

	Very low	Low	Average	High	Very High	Don't know
Satisfaction	9	8	2	0	0	0
Effort	9	10	0	0	0	0
Confidence	4	7	3	5	0	0
Cognitive Interest	7	8	2	0	0	2
Sensory Interest	11	5	0	1	0	2

Table D.2: Results of motivational model question, Simulation 2.

SITUATION 1

	Very low	Low	Average	High	Very High	Don't know
Satisfaction	0	0	6	9	4	1
Effort	0	1	14	3	1	1
Confidence	0	0	3	11	6	0
Cognitive Interest	0	2	12	2	3	1
Sensory Interest	0	1	10	3	3	3

SITUATION 2

	Very low	Low	Average	High	Very High	Don't know
Satisfaction	0	5	8	3	3	1
Effort	0	0	5	14	1	0
Confidence	0	6	5	8	0	1
Cognitive Interest	0	1	7	8	3	1
Sensory Interest	0	1	10	3	0	6

SITUATION 3

	Very low	Low	Average	High	Very High	Don't know
Satisfaction	0	4	10	6	0	0
Effort	2	7	7	4	0	0
Confidence	0	1	6	8	4	1
Cognitive Interest	2	5	5	4	1	3
Sensory Interest	1	6	4	2	0	7

continued on next page

continued from previous page

SITUATION 4

	Very low	Low	Average	High	Very High	Don't know
Satisfaction	0	1	0	15	4	0
Effort	0	0	3	16	1	0
Confidence	0	1	5	12	2	0
Cognitive Interest	0	0	2	14	4	0
Sensory Interest	0	0	9	5	0	6

SITUATION 5

	Very low	Low	Average	High	Very High	Don't know
Satisfaction	0	1	10	7	1	1
Effort	0	0	10	8	2	0
Confidence	0	0	8	11	1	0
Cognitive Interest	0	3	7	7	1	2
Sensory Interest	0	4	8	2	0	6

SITUATION 6

	Very low	Low	Average	High	Very High	Don't know
Satisfaction	0	0	0	6	14	0
Effort	0	0	1	10	9	0
Confidence	0	0	0	8	12	0
Cognitive Interest	0	0	2	5	11	2
Sensory Interest	0	2	4	7	5	2

D.4 Importance question results

We can see a complete list of all the results for the question regarding the importance of the motivational factors in the questionnaire (second question) in tables D.3 and D.4.

Table D.3: Results of importance question, Simulation 1.

SITUATION 1

	Not imp.	Slightly imp.	Important	Very imp.	Essential	Don't know
Satisfaction	0	0	6	11	3	0
Effort	0	0	5	12	3	0
Confidence	0	1	3	10	6	0
Cogn. Interest	1	1	1	12	5	0
Sensory Interest	0	1	8	9	2	0

SITUATION 2

	Not imp.	Slightly imp.	Important	Very imp.	Essential	Don't know
Satisfaction	0	0	7	9	3	0
Effort	0	1	4	10	3	1
Confidence	0	0	3	13	3	0
Cogn. Interest	0	1	3	11	3	1
Sensory Interest	0	1	4	9	3	2

SITUATION 3

	Not imp.	Slightly imp.	Important	Very imp.	Essential	Don't know
Satisfaction	0	2	7	7	3	0
Effort	0	1	4	8	6	0
Confidence	0	1	5	8	5	0
Cogn. Interest	0	2	5	5	7	0
Sensory Interest	0	2	4	5	6	2

continued on next page

continued from previous page

SITUATION 4

	Not imp.	Slightly imp.	Important	Very imp.	Essential	Don't know
Satisfaction	0	2	6	8	2	1
Effort	0	1	4	11	3	0
Confidence	0	0	4	11	4	0
Cogn. Interest	0	2	8	5	4	0
Sensory Interest	0	2	6	5	4	2

SITUATION 5

	Not imp.	Slightly imp.	Important	Very imp.	Essential	Don't know
Satisfaction	0	0	4	7	8	0
Effort	0	1	4	7	6	1
Confidence	0	1	3	9	6	0
Cogn. Interest	1	1	2	11	4	0
Sensory Interest	0	2	3	6	6	2

SITUATION 6

	Not imp.	Slightly imp.	Important	Very imp.	Essential	Don't know
Satisfaction	1	1	2	10	5	0
Effort	1	1	5	9	3	0
Confidence	1	1	4	7	6	0
Cogn. Interest	1	0	4	7	7	0
Sensory Interest	1	1	1	7	7	2

Table D.4: Results of importance question, Simulation 2.

SITUATION 1

	Not imp.	Slightly imp.	Important	Very imp.	Essential	Don't know
Satisfaction	0	2	10	4	3	1
Effort	0	1	8	7	2	2
Confidence	0	2	7	8	2	1
Cogn. Interest	0	0	7	8	5	0
Sensory Interest	1	5	6	6	2	0

SITUATION 2

	Not imp.	Slightly imp.	Important	Very imp.	Essential	Don't know
Satisfaction	0	1	6	8	5	0
Effort	0	2	9	7	2	0
Confidence	0	3	6	9	2	0
Cogn. Interest	0	0	7	10	2	1
Sensory Interest	0	5	7	4	0	4

SITUATION 3

	Not imp.	Slightly imp.	Important	Very imp.	Essential	Don't know
Satisfaction	0	1	7	9	3	0
Effort	0	1	5	9	4	1
Confidence	0	2	5	12	1	0
Cogn. Interest	0	0	4	12	3	1
Sensory Interest	1	4	4	8	0	3

continued on next page

continued from previous page

SITUATION 4

	Not imp.	Slightly imp.	Important	Very imp.	Essential	Don't know
Satisfaction	0	1	7	9	3	0
Effort	0	1	5	10	4	0
Confidence	0	1	7	10	2	0
Cogn. Interest	0	0	4	12	4	0
Sensory Interest	0	5	5	6	0	4

SITUATION 5

	Not imp.	Slightly imp.	Important	Very imp.	Essential	Don't know
Satisfaction	0	0	6	9	4	1
Effort	0	1	8	8	3	0
Confidence	1	2	5	11	1	0
Cogn. Interest	0	0	6	11	2	1
Sensory Interest	0	6	2	8	0	4

SITUATION 6

	Not imp.	Slightly imp.	Important	Very imp.	Essential	Don't know
Satisfaction	0	0	7	8	4	1
Effort	0	2	3	13	2	0
Confidence	0	3	8	9	0	0
Cogn. Interest	0	1	4	11	3	1
Sensory Interest	0	5	6	5	0	4

Appendix E

Published papers

The following papers have been published in connection with the research described in this dissertation. They are reproduced here with permission.

de Vicente, A. and Pain, H. (1998). Motivation diagnosis in intelligent tutoring systems. In Goettl, B. P., Halff, H. M., Redfield, C. L., and Shute, V. J., editors, *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, volume 1452 of *Lecture Notes in Computer Science*, pages 86–95, Berlin. Heidelberg. Springer. .... **page 317**

de Vicente, A., Bouwer, A., and Pain, H. (1999). Initial impressions on using the DISCOUNT scheme. In *Proceedings of the Workshop on Analysing Educational Dialogue Interactions: Towards Models that Support Learning (AI-ED'99)*, pages 87–94. .... **page 327**

de Vicente, A. and Pain, H. (1999a). Eliciting motivation diagnosis knowledge. In Lajoie, S. P. and Vivet, M., editors, *Proceedings of the Ninth World Conference on Artificial Intelligence in Education*, pages 651–653, Amsterdam. IOS Press. .... **page 335**

de Vicente, A. and Pain, H. (1999b). Motivation self-report in ITS. In Lajoie, S. P. and Vivet, M., editors, *Proceedings of the Ninth World Conference on Artificial Intelligence in Education*, pages 648–650, Amsterdam. IOS Press. ...... **page 338** de Vicente, A. and Pain, H. (1999c). Towards 'motivating' instructional systems. In G. Cumming, T. Okamoto, and G. Louis, editors, *Proceedings of the Seventh International Conference on Computers in Education*, volume 1, pages 744–751, Amsterdam. IOS Press. . . . . **page 341**

de Vicente, A. and Pain, H. (2000). A computational model of affective educational dialogues. In *AAAI Fall Symposium: Building Dialogue Systems for Tutorial Applications, North Falmouth, Massachusetts*, pages 113–121, Menlo Park, CA, USA. AAAI Press. Technical Report FS-00-01. . . . . **page 349**

de Vicente, A. and Pain, H. (2002). Informing the detection of the students' motivational state: An empirical study. In S. A. Cerri, G. Gouardères, and F. Paraguaçu, editors, *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems*, volume 2363 of *Lecture Notes in Computer Science*, pages 933–943, Berlin. Heidelberg. Springer. (*ITS2002 Best Paper Award*). . . . . **page 358**

Motivation Diagnosis in Intelligent Tutoring Systems

Angel de Vicente* and Helen Pain

Department of Artificial Intelligence, The Univerisity of Edinburgh
{angelv,helen}@dai.ed.ac.uk

Abstract. Despite being of crucial importance in Education, the issue of motivation has been only very recently explicitly addressed in Intelligent Tutoring Systems (ITS). In the few studies done, the main focus has been on motivational planning (i.e. how to plan the instruction in order to motivate the student). In this paper we argue that motivation diagnosis (i.e. how to detect the student's motivational state) is of crucial importance for creating 'motivating' ITSs, and that more research is needed in this area. After an introduction, we review some relevant research on motivation diagnosis, and then we suggest directions which further research in this area might take. Although the issues discussed here are still poorly understood, this paper attempts to encourage research in the ITS community in what we believe is one of the most important aspects of instruction.

1 Introduction

Students' motivation to learn is an important issue in Education and a basic concern in classroom practice. As Goleman puts it, "The extent to which emotional upsets can interfere with mental life is no news to teachers. Students who are anxious, angry, or depressed don't learn; people who are caught in these states do not take in information efficiently or deal with it well." [1, pp. 78].

But what do we mean by motivation? The literature about this subject is extensive, and it is outside the scope of this paper to review the many theories that have been proposed to explain motivation.[1] Here, the definition given by Williams and Burden, which fits well with the meaning of motivation intended in this paper, will suffice:

> *Motivation* may be construed as
> - a state of cognitive and emotional arousal,
> - which leads to a conscious decision to act, and
> - which gives rise to a period of sustained intellectual and/or physical effort

* Supported by grant PG 94 30660835 from the "Programa de Becas de Formación de Profesorado y Personal Investigador en el Extranjero", Ministerio de Educación y Cultura, SPAIN.

[1] For a good introductory book to the subject see [2].

– in order to attain a previously set goal (or goals).

[3, pp. 120], (emphasis in original)

Therefore, one of the main concerns in Education is how to induce that *state of cognitive and emotional arousal* which will make the instruction an interesting and engaging experience for the student. Thus, for instance, Lepper *et al.* report after studying the motivational techniques of expert tutors in elementary school mathematics activities that "Expert human tutors, it would appear, devote at least as much time and attention to the achievement of affective and motivational goals in tutoring, as they do to the achievement of the sorts of cognitive and informational goals that dominate and characterize traditional computer-based tutors." [4, pp. 99].

Consequently, it is surprising to find that very little research has dealt explicitly with motivational aspects of instruction in ITSs. It is true that many ITSs attempt to motivate the student by using multimedia, games, etc., but this approach seems to be based on the idea that it is possible to create instruction that is motivating *per se*. We believe, on the other hand, that ITSs would benefit from being able to adapt their instruction not only according to the student's cognitive state, but also to her motivational state.

This approach has been taken on very few occasions (e.g. [5, 6]). From these, the most comprehensive work is that of del Soldato [5], who added two new modules to the traditional ITS architecture: a motivation modeller and a motivational planner. But these attempts have focused mainly on motivational planning (i.e. how to plan the instruction in order to motivate the student), rather than on motivation diagnosis (i.e. how to detect the student's motivational state). In this paper we focus on the latter.

A first concern (and the issue around which this paper is centered) is the selection of a 'communication channel' that provides the necessary information to perform the motivation diagnosis. In the case of humans the available communication channels are many (verbal communication, the eyes, the hands, posture, body rhythms, smell, touch, etc.) [7], but not all of them are always needed, and we can communicate more or less effectively using only a number of them (for instance by telephone, letter, sign language, etc). Some advantages and disadvantages of using some of these communication channels in an ITS are given throughout the paper.

In the rest of this paper we explore some of these issues: human motivation diagnosis in section 2, some of the relevant research concerning motivation diagnosis in ITSs in section 3, and lastly we present some research suggestions in section 4.

2 Human Diagnosis of Motivation

Ironically, despite the large volume of research on human motivation, we still don't understand how we diagnose other people's motivation (or more generally, other people's emotions[2]). Picard even raises the question: "How do we detect a person's emotions? Is it via some metaphysical sixth sense?" [8, pp. 4-5].

In most cases, this ability to read other people's emotions—empathy—is taken for granted, although not everybody shows the same proficiency at it. A test for measuring this ability to empathise with others has been used by Robert Rosenthal and his students. The test—or Profile of Nonverbal Sensitivity (PONS)—consists of a series of videotapes of a young woman expressing different feelings (having the words muffled), which are shown to the viewers after blanking out one or more channels of nonverbal communication [9, cited in [1]].

What seems to be undisputed is that emotion diagnosis takes place mainly through non-verbal cues. For instance, Lepper *et al.* point out that although the analyses of the tutoring sessions were not completed, "an initial hypothesis is that our tutors' affective diagnoses depend much more heavily than their cognitive assessments on inferences drawn from the student's facial expressions, body language, intonation, and other paralinguistic cues." [4, pp. 101].

An ITS would also benefit by having this ability to empathise, but observations of human empathy makes us reflect upon the real difficulty of this ability. Goleman has argued that the ability to empathise "builds on self-awareness; the more open we are to our own emotions, the more skilled we will be in reading feelings." [1, pp. 96], and points out that alexithymics (people who lack words for their feelings[3]) lack the ability to 'read' other people's emotions. These observations raise the question of whether a computer should *have* emotions in order to be able to empathise with its user.

3 Computer Diagnosis of Motivation

One of the first suggestions of endowing computer tutors with a degree of empathy was made by Lepper and Chabay [10]. They argued that motivational components are as important as cognitive components in tutoring strategies, and that important benefits would arise from considering techniques to create computer tutors that have an ability to empathise.

More recently, Issroff and del Soldato [11] have suggested that in computer-supported collaborative learning settings, the instructional planner of a system should also include the goal of motivating the learner.

Although in both cases the main focus has been on motivational planning, Lepper and Chabay [10] also suggested some additions that should be made to a computer tutor in order to provide it with an ability to detect a student's motivational state: 1) make available to the computer tutor specific background

[2] In this paper motivation diagnosis is considered, for simplicity, as a particular case of emotion diagnosis.

[3] "Indeed, they seem to lack feelings altogether, although this may actually be because of their inability to *express* emotion rather than from an absence of emotion altogether." (pp. 50).

knowledge about the student (gathered from aptitude tests, motivational measurement scales or even teachers' assessments) and 2) enable the computer to ask the student directly whether she would like help or not, whether she would prefer harder or easier problems, etc.

In the following sections we review the research done on computer motivation diagnosis, classifying it according to the type of information used for the diagnosis.

3.1 Questionnaires

Questionnaires have been sometimes used for collecting information about the student's motivation to learn. Thus, Arshad [12, cited in [5]], used questionnaires applied at the beginning of the first interaction in order to model the student's confidence state[4].

Matsubara and Nagamachi [6] also used questionnaires[5] at the beginning of the interaction to diagnose several factors influencing motivation, such as: Achievement Motive, Creativity, Sensation Seeking Scale, Extroversion Intention, Work Importance and Centrality, Least Preferred Co-worker and Locus of Control.

Whitelock and Scanlon used post test questionnaires (consisting of five open ended questions) to assess a number of motivational factors, such as "curiosity, interest, tiredness, boredom and expectation plus the challenge of the task and partnership itself" [13, pp. 276].

Pre-interaction questionnaires have been criticised as being static, while the student's motivational state is likely to change during the interaction [5]. On the other hand, questionnaires can be useful means for detecting motivational traits (more enduring characteristics of students) and several tests have been devised to measure motivation. For instance, Gardner [14] devised an Attitude/Motivation Test Battery (AMTB), which consists of a series of self-report questionnaires, in order to calculate an Attitude Motivation Index (AMI), and O'Bryen [15] gives an introduction to the use and development of questionnaires to assess motivation in second language classrooms.

Therefore, we could use questionnaires for collecting information about enduring characteristics of the student that can help to adapt the instruction, although other methods should be used to gather information about more transient characteristics of the student's motivation.

3.2 Verbal Communication

The system by del Soldato [5] detected information about the motivational state of the student through three main sources: direct communication with the student during the interaction, perseverance to complete the task and student's requests for help.

[4] One of the many factors that influence motivation.

[5] But no precise account or reference to how the questionnaires were created or what they 'look like' is given in the paper.

Direct communication concerns the interaction and the task at hand, but not the motivational state of the student, which is inferred from the student's expressions. For example, certain answers are considered typical of low confidence students (e.g. "no, too difficult"), which would help the computer tutor to infer that the confidence of the student is low.[6]

The motivation diagnosis knowledge is implemented as a set of production rules that help the system 'decide' how to update the motivation model of the student, which is represented as a set of three numerical variables: effort, confidence and independence. For example, one of the rules indicates that the value of the confidence variable should be increased by a large constant if the student has performed a problem correctly without help.

This approach has the advantage of being relatively simple to implement, but since most of an emotional message seems to be non-verbal, it may prove to be too restricted for diagnosing student's motivation and it may be very difficult to elicit motivation diagnosis knowledge for this type of 'communication channel'. del Soldato [5] performed a preliminary study to test the accuracy of her motivational modeller, but no conclusive results were obtained, and therefore more research is needed.

3.3 Self-report

Another approach to motivation diagnosis is to have direct communication with the student about her motivational state.

This approach was taken by del Soldato [5] for the design of the system MORE, but we learn that, unfortunately, the features needed to read the self-report of the student's motivational state were not operational in the first version of the system, and hence we do not have an evaluation of how effective these features can be.

It is not clear whether or how the student's motivational state will be affected by having to report about her own motivation. But, if it is not affected significantly, this could be one of the easiest ways to diagnose it. As Briggs *et al.* put it, "confidence judgments are extremely simple to elicit, since users can give subjective ratings more easily than they can offer explanations." [16]

Different theories and models identify several factors that affect motivation. The ARCS Model by Keller [17], for instance, defines four components that influence the motivation to learn: attention, relevance, confidence and satisfaction. An interface could be easily implemented with mechanisms that would allow the student to report her subjective reading of these factors. For example, each of these factors could be represented by a slider, which could be manipulated by the student.

[6] The system lacks a natural language interface. Instead, the student selects her answer from a set of standard expressions.

3.4 Expert System

Hioe and Campbell [18] were concerned with performance problems in the workplace, and they devised a prototype expert system to help the employee's manager or supervisor to find which problems were affecting the employee's performance, with special emphasis on motivational problems. By reviewing different theories of human motivation, and based on the expert's diagnostic processes[7], they classified potential motivation problems into four groups: a) performance standards and goals; b) positive and negative outcomes; c) human relation issues; d) work itself. Then, they created a set of questions for specific motivational problems in these four groups, and developed an expert system that could ascertain which of these motivational conditions was causing the poor performance.

Although their work is not focused on computer tutors, we can imagine a similar approach being used in an ITS. This approach would not be as dynamic as other approaches (e.g. self-report), but neither as static as questionnaires. For instance, if at some point during the interaction a performance problem is found, the expert system could be used to directly ask the student in order to find the causes of the problem.

3.5 Sentic Modulation

Sentic modulation refers to "the physical means by which an emotional state is typically expressed" [19, pp. 25], and although still in its infancy, there has been some research addressing the issue of how a computer system could use this information to detect human emotions.

This approach has not been used, to our knowledge, in computer instruction systems, but Picard [8] reports the interest in the MIT Media Lab of building a piano-teaching computer system that could detect student's expressive timing.

In the group lead by her, the Affective Computing group, the research on detection of emotions is based primarily on the detection of patterns in physiological states. Thus, one of the areas of their research is to design sensors that will help to detect the user's emotional state. A Prototype Sensing System was developed, which includes four sensors: Galvanic Skin Response (GSR) Sensor, Blood Volume Pulse (BVP) sensor, Respiration sensor and Electromyogram (EMG) sensor.

Elias Vyzas has used this system in experiments of emotion recognition, obtaining in the best cases a 83% accuracy discriminating the physiological signals from an actress expressing different emotions [19, described in pp. 185-188].

Some work has also been done in recognising emotions from facial expressions in video, and from vocal intonation. Essa and Pentland [20, cited in [19]] developed a facial recogniser that showed an accuracy of 98% in recognising (although not in real time) six deliberately made facial expressions for a group of eight people. Roy and Pentland [21, cited in [19]] studied the possibility for a

[7] The expert being the second author.

computer of discriminating between approving or disapproving sentences, obtaining similar classification accuracy to humans, 65% to 88% for speaker-dependent, text-independent classification.

These methods could be used to detect the motivational state of a student in an ITS, but the approach seems much harder than some of the methods discussed in previous sections. On the other hand, we have seen that most of an emotional message is conveyed through non-verbal communication, and therefore, the use of a sentic modulation approach may prove to be the most efficient way to diagnose motivation.

In addition to the problem of efficiency of motivation diagnosis, we should also consider the difficulty of applying these methods to current ITSs, and in the case of physiological data, the possibly negative reaction of students to the use of body sensors. It could happen that despite being a very efficient way of detecting human emotions, its use could be limited to a certain type of application, as a consequence of users' reaction.

4 Research Areas Worth Pursuing

In such a new and complex area as motivation diagnosis, the list of possible research directions is extensive. Therefore, in this section we will simply present a few of the main issues that should be further explored.

Exploring Other Communication Channels. The research reviewed in this paper makes use of many different 'communication channels', but there are many other possibilities that could be studied. For example, a computer could determine the user's pupils size, which changes with different emotional states [22, cited in [7]], or could record mouse movements and typing speed and errors, or even could analyse the posture of all or some part of the body.

Eliciting Motivation Diagnosis Knowledge. In most cases, independently of the 'communication channel' that we use, the main problem is to know which pattern corresponds to which motivational state. Knowledge about motivation diagnosis may be elicited based on theories of motivation, observations of human teachers or even 'common sense', but we must devise experiments to test the validity of this knowledge.

Previously we mentioned one of the production rules used by del Soldato [5] to model student's confidence: '*increment by a large constant the confidence value if the student has succeeded solving a problem without help*'. This rule seems reasonable based on observations of human teachers, but the information available to humans is far more than the information available to the computer system, and we should study whether this knowledge is sufficient to detect the student's confidence level, or whether other factors are also influencing the decision.

Thus, we could devise experiments where humans should try to diagnose a student's motivational state by inspecting an instruction interaction where most of the 'communication channels' are blanked out (similar to the experiments mentioned in section 2 by Rosenthal *et al.*), maintaining only those channels available to the computer system (for instance, in the case of the production rule mentioned, information about the failure or success in performing the problem and information about help requests). With this type of experiments, we could start comparing the accuracy of different formalisations of motivation diagnosis knowledge.

Investigating Self-report Approaches. As we have seen, an approach to motivation diagnosis based on student's self-report would probably be the easiest to implement, but there are also many issues that should be studied. In section 3.3 we have mentioned, as an example, the use of sliders to represent each of the important factors that affect motivation. We should further investigate the effect on the user of using this approach: Does the student's motivational state change by having to inform about it? Does the student updates the sliders regularly, or only when certain 'extreme' events happen? Does she find the need to update the sliders regularly intrusive? We should also consider different ways of self-reporting motivational states: for example, we could develop an interface where several 'emoticons' are displayed, representing different states (e.g.: bored, anxious, happy,...), from which the user could select according to her own motivational state.

Developing Models of Motivation. Another interesting research direction that should be explored is the development of predictive models of motivation and their use as aids for diagnosis. Picard [19] describes some models of affective behaviour, but these have been mainly used for emotion synthesis.

The addition of these type of models could prove vital for the development of truly motivating computer instruction. Most probably, human tutors do not only react to motivational states, but make use of models of motivation to *predict* the motivational effect of possible actions, and thus to select a line of action that brings about the best possible outcome.

Exploring Individualised Motivation Diagnosis and Models of Motivation. We should consider the possibility of creating *individualised* methods of motivation diagnosis and models of motivation. del Soldato [5] considered, for simplicity, that the behaviour pattern of every student in a certain motivational state is basically the same. Similarly, we could implement general models of motivation to predict motivation outcomes for all students. But a good human tutor surely knows that each student has her own characteristics, and that different students can react differently to the same tutor actions.

As we approach the goal of creating computers systems that can detect our motivational states and even create a model of our motivational behaviour, we also face several dilemmas: Who could have access to this information? Where should we place the division between public and private emotional data? What would the reaction of the users be? It could be argued that given the state of research in this area, these questions are probably not very important at present. But the answer to these questions could give us important hints of whether or up to what point this type of research is worth pursuing.

For example, in an informal survey among 10 students in our Department we found—perhaps not surprisingly—that the use of physiological data was considered the most intrusive approach, and the one that they would least willingly use. On the other hand, the monitoring of keyboard and mouse actions was considered the least intrusive, and the approach that they would prefer to use. Although a formal study is needed, this seems to indicate that intrusiveness of the system would be an important factor when deciding whether to use it or not, and therefore it should be considered when investigating different approaches to motivation diagnosis.

5 Conclusion

In this paper we have focused on the issue of motivation diagnosis, which we believe to be of crucial importance for the development of truly effective ITSs.

We have reviewed different approaches taken for computer motivation diagnosis (or more generally emotion detection), finding that none of these approaches seem to offer a definitive solution. Therefore, we have suggested several areas of research that should be further explored.

While the efficient detection of a student's motivational state by an ITS seems to be a distant and difficult goal, we believe that research in this area can bring great benefits for computer instruction, and we hope that the issues presented in this paper are studied by the ITS community in order to make computer instruction more 'human'.

Acknowledgements

We thank Anders Bouwer, Jon Ander Elorriaga, Chi-Chiang Shei and anonymous reviewers for useful comments and suggestions on previous versions of this paper.

References

1. Goleman, D. *Emotional intelligence: Why it can matter more than IQ.* Boomsbury, London, 1996. First published by Bantam Books, New York, 1995.
2. Weiner, B. *Human motivation: metaphors, theories, and research.* Sage Publications Inc., 1992.
3. Williams, M. and Burden, R. L. *Psychology for language teachers: A social constructivist approach.* Cambridge University Press, Cambridge, UK, 1997.
4. Lepper, M., Woolverton, M., Mumme, D., and Gurtner, J. Motivational techniques of expert human tutors: lessons for the design of computer-based tutors. In S.P. Lajoie and S.J. Derry, editors, *Computers as cognitive tools.* Lawrence Erlbaum, 1993.

5. del Soldato, T. Motivation in tutoring systems. Tech. Rep. CSRP 303, School of Cognitive and Computing Sciences, The University of Sussex, UK, 1994.
6. Matsubara, Y. and Nagamachi, M. Motivation system and human model for intelligent tutoring. In Claude Frasson, Gilles Gauthier, and Alan Lesgold, editors, *Proceedings of the Third International Conference in Intelligent Tutoring Systems*, pages 139–147. Springer-Verlag, 1996.
7. Davis, F. *La comunicación no verbal*, volume 616 of *El Libro de Bolsillo*. Alianza Editorial, Madrid, Spain, 1976. Translated by Lita Mourglier from: "Inside Intuition - What we Knew About Non-Verbal Communication"; McGraw-Hill Book Company, New York.
8. Picard, R. W. Affective computing. Technical Report 321, M.I.T. Media Laboratory Perceptual Computing Section, Cambridge, Massachussetts, November 1995. Report available on-line on November 19, 1997 at http://vismod.www.media.mit.edu/cgi-bin/tr_pagemaker.
9. Rosenthal, R. *et. al.* The PONS test: Measuring sensitivity to nonverbal cues. In P. McReynolds, editor, *Advances in Psychological Assessment*. Jossey-Bass, San Francisco, 1977.
10. Lepper, M. R. and Chabay, R. W. Socializing the intelligent tutor: Bringing empathy to computer tutors. In Heinz Mandl and Alan Lesgold, editors, *Learning Issues for Intelligent Tutoring Systems*, chapter 10, pages 242–257. Springer-Verlag, 1988.
11. Issroff, K. and del Soldato, T. Incorporating motivation into computer-supported collaborative learning. In Brna et al. [23], pages 284–290.
12. Arshad, F. *The Design of Knowledge-based Advisors for Learning*. PhD thesis, School of Education, University of Leeds, UK, 1990.
13. Whitelock, D. and Scanlon, E. Motivation, media and motion: Reviewing a computer supported collaborative learning experience. In Brna et al. [23], pages 276–283.
14. Gardner, R. C. *Social psychology and second language learning: the role of attitudes and motivation*. London: Edward Arnold, 1985.
15. O'Bryen, P. Using questionnaires to assess motivation in second language classrooms. *University of Hawaií Working Papers in ESL*, 14(2), 1996.
16. Briggs, P., Burford, B., and Dracup, C. Self confidence as an issue for user-modeling. In *Proceedings of the Fifth International Conference on User Modeling*, Kailua-Kona, Hawaii, 1996. User Modeling, Inc.
17. Keller, J. M. Strategies for stimulating the motivation to learn. *Performance and Instruction Journal*, 26:1–7, 1987.
18. Hioe, W. and Campbell, D. J. An expert system for the diagnosis of motivation-related job performance problems: Initial efforts. DICS publication no. TRA3/88, Department of Information Systems and Computer Science, National University of Singapore, 1988.
19. Picard, R. W. *Affective Computing*. The MIT Press, Cambridge, Massachusetts, 1997.
20. Essa, I. and Pentland, A. Coding, analysis, interpretation and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763, July 1997.
21. Roy, D. and Pentland, A. Automatic spoken affect analysis and classification. In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, pages 363–367, Killington, VT, Oct 1996.
22. Eckhard H., H. Attitude and pupil size. *Scientific American*, 212:46–54, 1965.
23. Brna, P., Paiva, A., and Self, J., editors *Proceedings of the European Conference on Artificial Intelligence in Education*, 1996.

Initial impressions on using the DISCOUNT scheme*

Angel de Vicente
angelv@dai.ed.ac.uk
Division of Informatics, Artificial Intelligence University of Edinburgh

Anders Bouwer
anders@swi.psy.uva.nl
Social Science Informatics Department of Psychology University of Amsterdam

Helen Pain
helen@dai.ed.ac.uk
Division of Informatics, Artificial Intelligence University of Edinburgh

Abstract

In this paper we present some of our first impressions on using the DISCOUNT marking scheme. Our main objective was to study the reliability of the scheme and its usefulness to answer questions relevant to our research. Thus, three educational dialogues were marked-up and analysed using the DISCOUNT scheme by two of us in order to compare inter-coder reliability. In addition, we performed a simple analysis of the coded dialogues to study the research possibilities offered by the DISCOUNT scheme in conjunction with the analysis tool TACT. In general, the inter-coder reliability seems to us the main drawback, given the great number of mark-up codes available. But at the same time, its high flexibility seems to provide good opportunities for answering research questions relevant to educational systems.

1 Introduction

In the context of the workshop on "Analysing Educational Dialogue" at the University of Leeds, 1998, we wanted to study how the DISCOUNT scheme [Pilkington, 1997][1] could help us in our research on Educational Dialogues. Three educational dialogues were gathered, and we performed a preliminary analysis of these.

Our two main concerns were: 1) the inter-coder reliability and, 2) more importantly, the potential of the DISCOUNT scheme (in conjunction with the analysis tool mentioned in it, TACT [Lancashire, 1996]) in answering the type of research questions concerning educational dialogues that we were interested in. In the following sections we give a brief introduction to the DISCOUNT scheme, present the materials and tools used for our study, and then we summarise the analysis results. We end up with a number of conclusions about the DISCOUNT scheme.

*This paper is based on the presentation given in the workshop "Analysing Educational Dialogue", University of Leeds, April 1998.

[1]The latest version of the scheme [Pilkington, 1999] has a number of changes, but the study mentioned here was performed using the previous version [Pilkington, 1997]. However, these changes do not affect the contents of this paper in a significant way.

2 The DISCOUNT Annotation Scheme

The DISCOUNT scheme was interesting to us because it reflected well some of our research interests, having been developed "to help describe and evaluate educational discourse and, to mark representational levels of discourse which might be necessary for the generation of natural dialogues by machine." [Pilkington, 1997]

A full description of the DISCOUNT scheme is outside the scope of this paper, so in this section we will simply explain the overall structure of the scheme, together with the procedure used for marking.

The overall structure of the scheme can be seen in figures 1 and 2. DISCOUNT has as its top level division of text what is called Episodes, which "consist of propositions on a topic linked by rhetorical relations in a developing focus-space." [Pilkington, 1999, pp. 12].

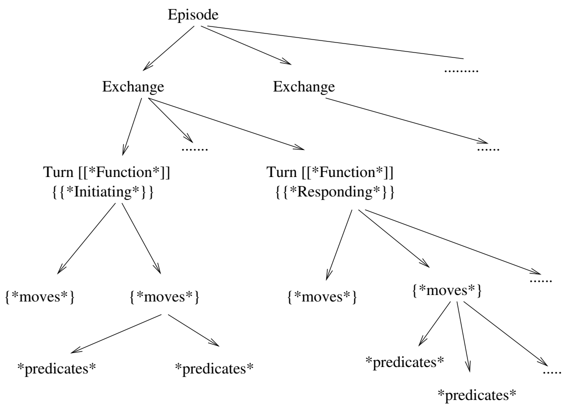


Figure 1: Coding Scheme Structure

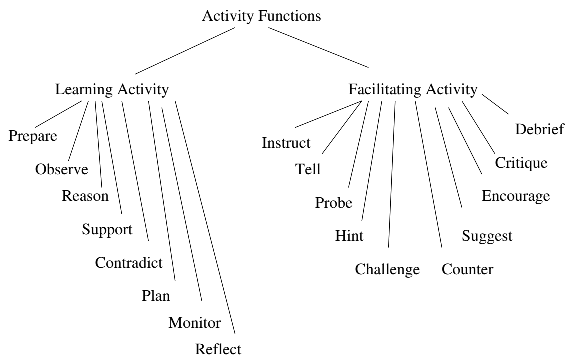


Figure 2: Activity Functions

As can be seen in figure 1, Episodes are divided in Exchanges, each of them consisting of Turns (initiating, responding or reinitiating). Each Exchange consists of at least one initiating and one responding Turn. At the same time, each Turn consists of Moves (propositions with a speech act function), and each Move has a number of Predicate labels (e.g. *Identify*, *Negate*, etc.). Each Turn can also have one or more Activity functions. The list of possible Activity functions is given in figure 2.

The procedure we used for marking consisted of approximately the following steps[2]:

1. Divide the transcript into episodes according to dialogue task and subtasks.

2. Categorise the topic status (if interested).

3. Categorise turn after turn on the move, dialogues role and predicate level:

    - segment turns by function, afterwards categorise as moves
    - moves assigned by pragmatic function in dialogue, usually clause or subclause
    - if unsure, find rhetorical predicates first

[2]Thanks to Cornelia Kneser for suggesting this procedure.

First turn of episode always to be an initiating one:

- each exchange starts with initiating and ends with responding role turn
- usually exchange only 2/3 turns and each utterance of 1 speaker is one turn
- re-initiating roles assigned if turn focuses on slightly different aspect

4. Mark up teaching/learning functions – as many as you want to each turn.

5. Mark the commitments and outcomes.

3 Analysis

For this study we analysed three educational dialogues[3]. Table 1 gives a brief description of the characteristics of each of the dialogues. We also include, as an example, the marked-up version of a portion of dialogue 2 as appendix A.

	Subject	Participants	Length (approx. number of words)
Dialogue 1	Euler's Circles Tutorial Dialogue	Tutor and student	2000
Dialogue 2	Tutorial dialogue on syntax tree construction	Tutor and student	1400
Dialogue 3	Medical Simulation Dialogue with experimenter (not tutor), and two students.	Experimenter and 2 students	7100

Table 1: Characteristics of dialogues analysed.

Each dialogue was analysed by one of us, with the exception of dialogue 2, which was analysed by two of us in order to compare inter-coder reliability.

The statistical analysis of the marked dialogues was performed with the software TACT for PC-compatible computers [Lancashire, 1996], which is also suggested in the DISCOUNT documentation, [Pilkington, 1997].

4 Results

4.1 Inter-scorer reliability

The inter-scorer reliability was, as expected, one of the main problems when using the DISCOUNT scheme. After having marked at least one dialogue in a previous session, and after

[3]The same dialogues were analysed by other participants of the workshop with other mark-up schemes to compare each scheme's strong and weak points.

having reached a reasonable level of agreement in the marking procedure (approximately 20 hours of experience of scheme usage), two of us marked a portion of dialogue 2.

	Marker 1	Marker 2
Total number of tags	119	85
identify	12	4
timesequence	1	7
analogy	-	3
conclusion	4	-
equivalent	5	-
settingtime	8	-
task	4	-
{*inform*}	13	5

Table 2: Main tagging differences

The main differences found can be seen in table 2. The first important difference is the number of tags used. As can be seen, Marker 1 used 119 tags while Marker 2 used only 85 for the same portion of the dialogue.

From the tags used in the marked dialogue, those whose frequency differed more greatly are also shown in table 2. All except one of them ({*inform*}) are predicates rather than moves. The frequency of higher-level tags was more similar in both versions of the marked dialogue, somehow to be expected, since the number of lower level tags is greater.

While we did not check for intra-scorer reliability, the differences found in our simple inter-scorer comparison are considerable, and we believe that an important effort should be made to ensure marking reliability (perhaps in terms of a thorough standardisation of the marking procedure).

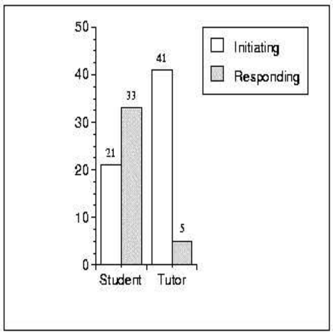
4.2 Research questions

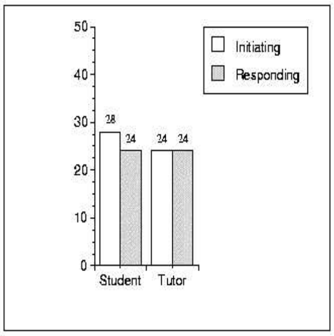
In this section we present the analysis performed for the given dialogues. Each of the following three subsections present a question related to some of our research. Even though the analysis was not very deep, it shows the type of results that could be obtained by simple dialogue analysis.

4.2.1 Who has the initiative? – Tutor vs. Student

Our first question was related to the initiative in the dialogue. Does the teacher always initiate a new exchange and does the student mainly respond, or does the student also initiate new exchanges?

To study this we looked at the tags corresponding to turns {{*initiating*}} and {{*responding*}}. The frequency of these tags in the three dialogues analysed can be seen in figures 3, 4 and 5. From these, it is clear that the dialogues differ greatly in relation to this issue.





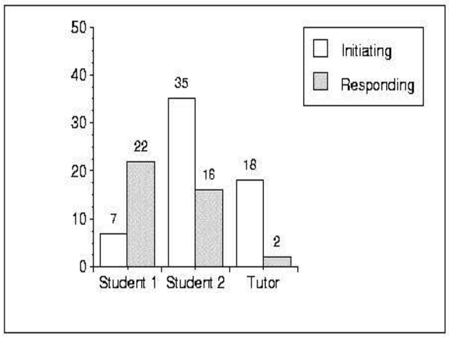


Figure 3: Dialogue 1 Figure 4: Dialogue 2 Figure 5: Dialogue 3

In dialogue 1 most of the initiating moves are done by the teacher, while she does virtually no responding. Dialogue 2 seems to present a more collaborative setting, where the student has even more initiative moves than the teacher, while the teacher has a high number of responding turns. Dialogue 3 seems to be a combination of the previous types. Student 2 presents most of the initiating moves, and Student 1 mainly responds while having basically no initiating moves.

Although the degree of participation of a student is not necessarily a strong indicator of effective learning, an instructional system could greatly benefit from this knowledge. In one-to-one tutoring, this is perhaps not a very useful indicator, unless it is also informed by the strategy used by the teacher, the task at hand, etc. But in a one-to-many situation it could be particularly useful.

For instance, we may want to have similar amounts of participation among students. While the task of judging the turn role of any particular utterance would be impossible for the majority of current systems, an interface similar to those used in web-based discussion groups could help to make explicit whether the student is responding or whether he is initiating a new exchange. This information could help the system to encourage or direct the students to participate, in order to obtain the desired participation distribution.

4.2.2 Reflection in Tutor-Student Dialogues

While the amount of initiating moves in a tutorial dialogue could be a signal for engaged and active students, it could also mean (as experienced regularly by any teacher) a student keen on participating, but not necessarily on the right track to effective learning.

In order to study this, we considered reflection as an alternative indicator of student's effective learning. As in the previous section, we operationalised the definition of reflection on the basis of several tags. In this case we looked principally at two types of tags. At the learning function level we looked for the tags [[*monitor*]] and [[*reflect*]], which are defined respectively as "Evaluate progress toward goal" and "Sum up a line of argument or revise reasoning in the light of new observations" [Pilkington, 1997]. Although these two techniques on their own are not sufficient condition for learning, they are in general considered good learning techniques.

For this question we compared dialogues 1 and 2 (both one-to-one settings). From table 3 it seems clear that the student in dialogue 2 is more 'reflective', at least according to our operationalised definition of the word. 37% of his learning function tags are either monitor or reflect as opposed to only 24% in the case of the student dialogue 1. Similarly 35% of the moves are metastatements compared to only 18% in dialogue 1. This role seems to be taken in dialogue 1 by the teacher, with a 38% of metastatements as opposed to only 13% in dialogue 2.

So, it seems clear that the student in dialogue 2 has better reflective abilities than the student in dialogue 1. When considering why this could be the case, we thought that the role of the teacher surely had to influence this, so we set out to study whether challenging the student will lead to him being more reflective. The occurrence of [[*challenge*]] or [[*counter*]] facilitating functions is also given in table 3.

	Dialogue 1	Dialogue 2
Learning Function: %Monitor/Reflect	24%	**37%**
Student Moves: %Metastatements	18%	**35%**
Tutor Moves: %Metastatements	**38%**	13%
Facilitating Functions: %Challenge/Counter	**28%**	10%

Table 3: Reflection in dialogues 1 and 2

It turns out that it is actually in dialogue 1 where the teacher challenges the student more. While this is against our expectations, it is certainly not conclusive. The tasks, the students, the teachers ... all were different between dialogues 1 and 2. But studying similar issues in dialogues where the settings are similar (same teacher and student, similar tasks, etc.) could help us to find out those teacher strategies that bring about desirable responses in the student.

4.2.3 Disagreement between participants

Our last question dealt with the disagreement between participants, and we looked at dialogue 3 (one experimenter and two students) for this. The predicates used as the operational definition of disagreement were *disagree*, *contraindication* and *maybenot*. The data for this dialogue is shown in table 4.

		with whom				
		S1	S2	Tutor	Anyone	% of total predicates
who disagrees	S1		2		2	9%
	S2	4	1	2	7	13%
	Tutor		1		1	7%

Table 4: Disagreement between participants

The rows indicate who disagrees, and the columns indicate who he/she disagrees with. As seen in section 4.2.1, student 2 has the highest initiative of the three participants, so it is not surprising to see that he is also the person who disagrees the most (he even disagrees with himself once).

Regarding the usefulness of this information for an instructional system, we can think of a setting where (similar to that outlined in section 4.2.1) the information about disagreements could be made explicit to the computer through special characteristics of the interface. This could be done by offering to the student different ways to interact with the system when he wants to disagree (perhaps indicating explicitly which previous utterance he is disagreeing with).

Again, this information could help to guide the instruction. For instance, a student who disagrees with himself very often could be regarded as too impulsive and temperamental; a student who disagrees with others too often (or perhaps with another participant in particular), could be regarded as too aggressive against others.

5 Conclusion

After our first experience using the DISCOUNT scheme we still had many questions about it, despite the long and detailed documentation in [Pilkington, 1997].

Many of the questions were related to particular aspects of the marking-up process, which stemmed from our desire to obtain a clear procedure of marking that would ensure a high inter-coder reliability. For instance, we were not sure whether we should try a top-down or bottom-up approach, or how to rationalise the mixed strategy that we used at the end.

Some of the categories were very difficult to code, particularly the dialogue roles. Similarly, sometimes it was difficult to decide which category to assign, and sometimes it was very tempting to invent a new category. Obviously, this would make the scheme ‘unstable’, which is probably not a good idea unless we find similar cases very frequently.

These and other questions lead us to the conclusion that inter-coder reliability would be the main drawback of the scheme, although this is a common problem for marking schemes [Carletta *et al.*, 1997]. Besides, DISCOUNT is a relatively large and complex scheme, and this leads to greater chances for coding variability. On the other hand, we believe that the scheme can become a very useful research tool. As we have seen in section 4, a number of research questions were easily approached after the coding of the dialogues.

Thanks to the variety of tags, it is possible to operationalise a great number of, otherwise, not very clear definitions. For instance, as seen in section 4.2.2, a complex concept such as reflection can be operationalised easily by defining it as the occurrence of a number of particular tags. How well a number of tags define a given concept is an issue for discussion, but selecting a fixed number of tags as its definition allows us to start performing useful comparisons between different dialogues. In general, while the task of marking a dialogue is certainly very time-consuming, we believe that the benefits obtained can be substantial.

A Example of marked-up dialogue

Episode

(*Tutor*) *Initiating* *Direct* *Action* *TimeSequence* *metastatement* *Self* *Other* *Now I’d like you to construct a tree for the sentence ’If Etta caught a bird she ate it’* [[*Instruct*]]

(*Student*) *Responding* *Accept* *Confirm* *OK. .Right*

Re-initiating *Metastatement* *Group* *TimeSequence* *so this time we’ve got if*

Metastatement *Self* *Reason* *Analogy* *Solution* *so it looks to me like I use... if and now I guess it'll be another of the basic sentences* [[*Observe*]] [[*Reason*]] [[*Plan*]]

(*Tutor*) *Responding* *Prompt* *Uhuh* [[*Encourage*]]

(*Student*) *Initiating* *Reason* *TimeSequence* *em..starting off with Etta again...*

Inform *Identify* *and now the verb in that was If Etta chased a bird*

Reason *Support* *so that would be the-.* [[*Observe*]] [[*Reason*]]

(*Tutor*) *Re-initiating* *Inform* *Identify* *If Etta caught a bird she ate it* [[*Tell*]]

(*Student*) *Responding* *Accept* *Confirm* *OK*

Reason *Analogy* *Support* *TimeSequence* *so that would be... sort of the same as the last one -*

Inform *Identify* *Reason* *Negate* *if Etta caught - now, a bird..... it's not a pronoun or a proper name*

Metastatement *Self* *Reason* *Select* *Instrument* *Achievement* *so I'll go for a determiner which will let me now do a.... bird.* [[*Observe*]] [[*Reason*]] [[*Plan*]]

Metastatement *Self* *Reason* *TimeSequence* *Achievement* *OK so now I have the first bit If Etta caught a bird ...em...*

Metastatement *Self* *Reason* *Analogy* *so I guess this'll be another simple sentence*

Reason *TimeSequence* *Self* *Inform* *Identify* *and this time it's a... I would guess it's a.... proper name again .....*

Inform *Contrast* *TimeSequence* *except that this time em, ah but the sentence, em, If Etta caught a bird she ate it,* [[*Monitor*]] [[*Plan*]] [[*Reflect*]]

Acknowledgements

The first author is supported by grant PG 94 30660835 from the "Programa de Becas de Formación de Profesorado y Personal Investigador en el Extranjero", Ministerio de Educación y Cultura, SPAIN. We also would like to thank Kaska Porayska-Pomsta for useful comments on previous versions of the paper.

References

Carletta, J., Isard, A., Isard, S., Kowtko, J. C., Doherty-Sneddon, G., and Anderson, A. H. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics*, **23**, 13–31.

Lancashire, I., editor (1996). *Using TACT with Electronic Texts*. The Modern Language Association of America, New York. Software available (on 30/4/99) at: ftp://chass.utoronto.ca/pub/cch/tact/tact2.1/.

Pilkington, R. M. (1997). Analyzing educational discourse: The DISCOUNT scheme. Technical report no. 019703, Computer Based Learning Unit, The University of Leeds, Leeds, UK.

Pilkington, R. M. (1999). Analyzing educational discourse: The DISCOUNT scheme. Technical report no. 99/2, Computer Based Learning Unit, The University of Leeds, Leeds, UK.

Eliciting motivation diagnosis knowledge

Angel de Vicente and Helen Pain
{angelv,helen}@dai.ed.ac.uk

Division of Informatics, Artificial Intelligence
The University of Edinburgh

1 Introduction

In a previous paper we emphasized the need of research in the area of motivation diagnosis in instructional systems, pointing out how little work had been done in the area [1]. While some of the previous work concerned with motivation in ITSs (e.g. [2]) dealt mainly with instructional planning, we focus in this paper on how to elicit and formalise knowledge to diagnose the student's motivational state while interacting with an instructional system.

The issue of how human teachers detect their students' motivation seems to be taken for granted, and it has been virtually unexplored in AI and Education research. Introspection and observational studies could throw some light into this issue, but they may be of limited usefulness for motivation diagnosis in ITSs. In a 'traditional' instructional setting or any other social interaction there is an 'overflow' of information. An incredible amount of information is available through various communication channels, such as facial cues, intonation, posture, etc. [3]. Many 'cues' that help us detect other people's emotions are perceived unconsciously via these channels, which makes it difficult to elicit emotion detection knowledge.

In order to limit the amount of sources of information available for knowledge elicitation, we designed a study in which a tutor will see exclusively the screen interaction of a student with an instructional system. That is, the tutor will be able to see in a computer screen only the interface of the instructional system which the student is manipulating.

We expect that it will be easier for tutors to rationalise their motivation diagnosis knowledge in this setting than if they were presented, say, with video-recordings of tutoring interactions. At the same time, we believe that the knowledge thus inferred will also be easier to formalise in terms of information available to the instructional system (such as time of interaction with system, mouse movements, etc.).

In the following sections we describe the planned study in greater detail and comment on the expected outcome of the study.

2 Description of study

In order to perform this study, we developed a prototype ITS called MOODS (MOtivation Diagnosis Study). MOODS was developed with two main functioning modes: student mode and teacher mode. In the student mode, MOODS is a simple tutoring system with an added facility that lets students inform the system about their motivational state. MOODS was used in this mode for a previous study [4].

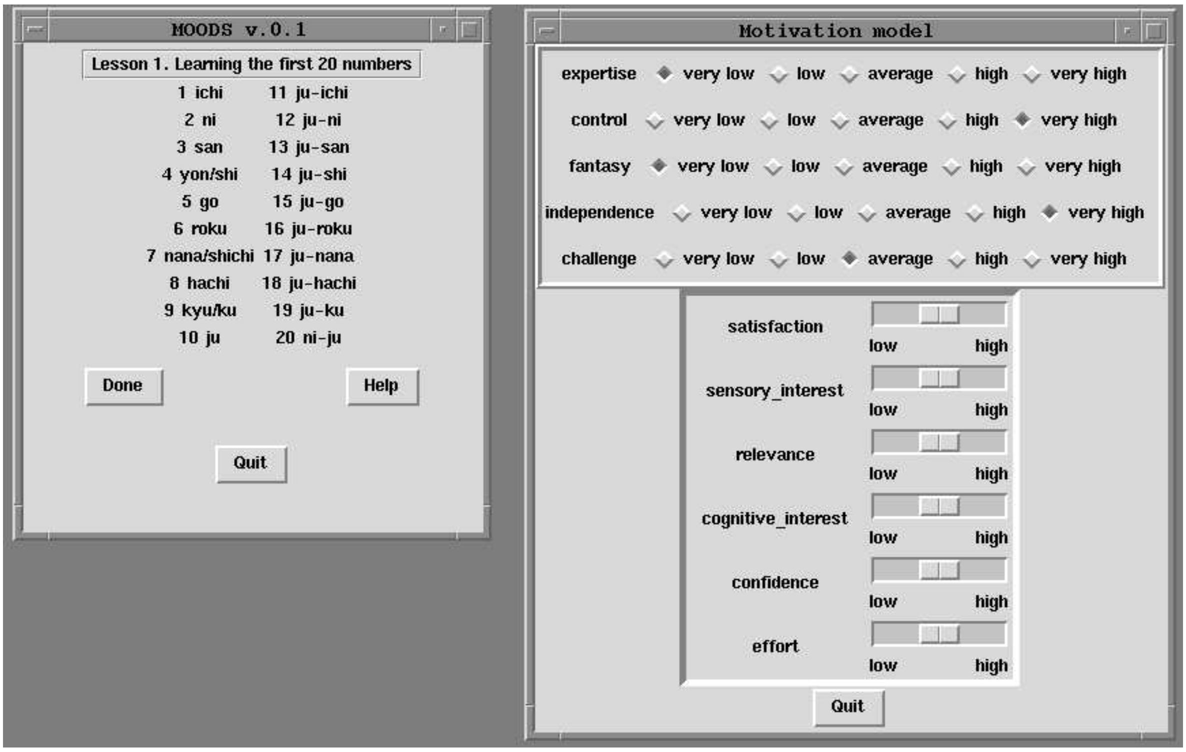


Figure 1: Teacher mode interface

In the teacher mode, MOODS can be used to see replayed the actions of a previous student interaction. The recording and replaying facilities were possible thanks to the program TkReplay [5]. The MOODS interface in this mode can be seen in figure 1.

The window to the left (MOODS v.0.1) is where the actions made by the student will be replayed[1]. The window to the right (Motivation model) is a representation of the motivational model of the student which is based mainly on two theories of motivation in education [6, 7], and which is described in more detail in [4]. The top part of this window shows certain traits ('permanent' characteristics) of the student whose interaction is being replayed. This information comes from previous student interactions with MOODS system. 18 such interactions were recorded as part of the study presented in [4]. The bottom part of this window will allow the participants of our study to predict the student's motivational 'state' at that time during the interaction. The interaction of the participants will be recorded again with the TkReplay software. Lastly, interviews with participants while showing them a replay of their own predictions will help us to rationalise their diagnosis of the student's motivational state.

3 Expected outcome

From this study we expect to learn about several issues, some of which are discussed in this section.

Is there a positive relation between self-perceived and tutor-inferred motivation data? As mentioned earlier, the MOODS interface used in the study presented in [4] has facilities that allow students to report about their motivational state as perceived by

[1]Certain aspects of the interaction, such as speed of replay, can be altered at will.

themselves. The interface for this was the same as the one which tutors are asked to update in this study (bottom part of right window in figure 1). This will allow an easy comparison between the self-perceived and tutor-inferred data.

How does 'trait' affect the motivation prediction? Tutors will have available in the interface a schematic representation of some of the student's traits (top part of right window in figure 1). How does this affect the prediction that tutors make about student's motivation? This could be useful for the implementation of instructional systems that base their instructional (motivational) planning according to student types.

Can we 'distill' general motivation diagnosis knowledge? From the tutors' interaction with the system and the interviews, we expect to elicit knowledge about motivation diagnosis. How does knowledge elicited from different tutors compare to each other? How does it relate to previous attempts of formalising motivation diagnosis knowledge [2]?

4 Conclusions

In this paper, we have presented a planned study to elicit motivation diagnosis knowledge. Despite the complex nature of motivational issues, we believe that this study can throw some light on how to design more effective instructional systems, and we look forward to present in due course the results of this study.

Acknowledgements

We would like to thank Chi-Chiang Shei, Marco Carvalho and anonymous reviewers for comments on previous versions of this paper.

The first author is supported by grant PG 94 30660835 from the "Programa de Becas de Formación de Profesorado y Personal Investigador en el Extranjero", Ministerio de Educación y Cultura, SPAIN.

References

[1] A. de Vicente and H. Pain. Motivation diagnosis in intelligent tutoring systems. In B. P. Goettl, H. M. Halff, C. L. Redfield, and V. J. Shute, editors, *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, pages 86–95, Berlin. Heidelberg, 1998. Springer.

[2] T. del Soldato. Motivation in tutoring systems. Tech. Rep. CSRP 303, School of Cognitive and Computing Sciences, The University of Sussex, UK, 1994.

[3] F. Davis. *La comunicación no verbal*, volume 616 of *El Libro de Bolsillo*. Alianza Editorial, Madrid, Spain, 1976. Translated by L. Mourglier from: "Inside Intuition - What we Knew About Non-Verbal Communication"; McGraw-Hill Book Company, New York.

[4] A. de Vicente and H. Pain. Motivation self-report in ITS. *In this volume.*

[5] C. Crowley. Tkreplay. version 1.1. Software downloadable at http://www.cs.unm.edu /~crowley, 1996. Available on-line on 18th November 1998.

[6] T. W. Malone and M. R. Lepper. Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow and M. J. Farr, editors, *Conative and Affective Process Analyses*, volume 3 of *Aptitude, Learning, and Instruction*, chapter 10, pages 223–253. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, 1987.

[7] J. M. Keller. Motivational design of instruction. In C. M. Reigeluth, editor, *Instructional-design theories and models: an overview of their current status*, pages 383–434. Lawrence Erlbaum Associates, 1983.

Motivation self-report in ITS

Angel de Vicente and Helen Pain
{angelv,helen}@dai.ed.ac.uk

Division of Informatics, Artificial Intelligence
The University of Edinburgh

1 Introduction

Motivation diagnosis is a virtually unexplored issue in AI and Education research, and we have previously emphasized the need for research in this area [1]. In this paper we present an empirical study in which self-report was used as a way of communicating with the computer about participants' motivation.

The self-report approach is probably one of the easiest to implement, and was suggested by [2] but, to our knowledge, has not been yet implemented in an ITS. To allow us to study students' reaction to and usefulness of the self-report approach, we developed a prototype ITS called MOODS (MOtivation Diagnosis Study) and performed an empirical study which we describe and discuss in the following sections.

2 Description of MOODS

MOODS was developed as a prototype of a simple tutoring system with an added facility that lets students inform the system about their motivational state as the instruction takes place. MOODS consists of 10 lessons, with the aim of teaching Japanese numbers up to 100.

Five types of lessons were developed ranging from simple presentation of numbers to memorise to a Tetris-like game to perform number additions. For each of these types we developed 2 lessons, one dealing with numbers up to 20 and another dealing with numbers up to 100.

We created six instructional paths, each of them consisting of a different combination of lessons. We attempted to create each instructional path with a different teaching style that would influence differently students' motivation. With this we hoped to see the reaction of students to the self-report method under a broad spectrum of affective situations.

As a central component of MOODS, we implemented a "motivation model" that is divided in two main categories: traits ('permanent' characteristics: independence, challenge, control, fantasy and expertise) and states ('transient' characteristics: effort, confidence, sensory interest, cognitive interest, relevance and satisfaction). This model is based primarily on two theories of motivation in education [3, 4] and on one of the very few motivational models implemented in an ITS to date [2].

Although the task of choosing between different factors is a difficult one, we believe that our model presents a useful set of the main important characteristics for student motivation, while doing it with a small number of variables which makes it feasible to use it as part of the self-report method.

Information about student's traits characteristics is gathered prior to the interaction with the main MOODS interface through a small on-line questionnaire in which the student has to rate his 'level' for each of the five traits categories. Information about student's motivational state is gathered throughout the interaction with MOODS via the manipulation by the student of six sliders, which represent the six 'state' categories of our motivation model.

3 Description of study and results

18 university students volunteered to participate in our study (3 students for each of the instructional paths). Their collaboration consisted of: a pre-questionnaire regarding the trait characteristics; interaction with MOODS system; and a post-questionnaire regarding mainly their opinions on the system and the usage of the motivational model.

The participants were asked to "use these sliders [representing the various motivational factors] **as often as possible** whenever you think there is a change in any of these factors, since it is necessary for the computer to understand your current situation in order to modify the instruction accordingly." (emphasis in original instructions). Actually the instructional paths are fixed, but this together with the true purpose of the study was explained to the participants only after they had filled in the post-questionnaire.

The interactions with MOODS (which on average lasted about 14.5 minutes) were recorded thanks to the software TkReplay [5]. Some preliminary results of the analysis of these data are presented below.

On acceptance of the self-report method. Participants' answers to the post-questionnaire seem to indicate that self-report could be an acceptable method for motivation diagnosis. The acceptance of the trait questionnaire was very high (an average answer of 4.21 in a range between 1 and 5). The acceptance of the motivational sliders (average answer was 4) was lower, although still very high given our expectations. Nonetheless, this acceptance may be artificially high due to the short length of the interaction with the system.

Which sliders were used more often? To inform future designs of self-report interfaces, it is important to understand which sliders were used more often, in order to give priority to these factors. The order in which the sliders were used was: confidence (with an average of 3.5 uses per participant), effort, satisfaction, sensory interest, cognitive interest and relevance (with an average of 1.33 uses per participant).

This is consistent with the answers to the post-questionnaire , in which participants were asked to comment on whether they found that any of the motivational model factors was particularly difficult to answer. Four people answered this, *relevance* being particularly difficult to answer for all four, *independence* by two of them, and each of *control, confidence* and *cognitive interest* by one of them.

Which values did the sliders take? From the point of view of designing a better self-report interface, it is also very important to know which values the sliders took during the interactions. It is also very important in order to understand the relations (if any) between different motivational factors.

Our hypothesis that some of the instructional paths would be more appealing than others seems to be confirmed for some of the factors (mainly *cognitive interest* and *sensory interest*, in which the value of these correlates to the number of instructional path), though not for others (mainly *relevance* and *effort*).

Also apparent is the positive relation between *confidence* and *satisfaction*. Our data only provides (on average) low levels of confidence, for which we would expect the observed positive relation. In future analysis we will analyse the data for individual lessons to see whether the relation between *confidence* and *satisfaction* is inverted, as expected, in those participants with a higher level of *confidence*.

4 Conclusion

In this paper we have presented a study of the possibility of use of self-report to detect motivation in an ITS. This is, to our knowledge, the first study of this type to date. As a result, we think that self-report is a viable option for motivation diagnosis. However, a number of issues have been raised and we intend to modify our approach for future implementations of MOODS.

As we suggested, the reaction of the system to the values of the motivational sliders may have a great impact on the willingness to continue using the self-report facilities. Therefore, care should be paid to make the reaction of the system to users updates of the sliders very obvious, in order to encourage the use of the self-report facilities.

As a continuation to this work, we will perform a study in which participants with tutoring experience will see previous interactions with MOODS replayed and will be asked to predict students' updates of the motivational model (see [6] for further details). That study will help us to find which variables from the interaction seem to be more important for the tutors in order to detect students' motivation. This could, in turn, help us modify the motivation model presented in section 2.

Acknowledgements

Thanks to Jeff Hobbs for letting us use and modify the code of his Tetris game (available at http://sunscript.sun.com/plugin/tetris.html on 18th November 1998), in order to turn it into an instructional game for MOODS.

We also would like to thank Kaska Porayska-Pomsta, Eshaa AlKhalifa, Marco Carvalho and anonymous reviewers for useful comments on previous versions of this paper.

The first author is supported by grant PG 94 30660835 from the "Programa de Becas de Formación de Profesorado y Personal Investigador en el Extranjero", Ministerio de Educación y Cultura, SPAIN.

References

[1] A. de Vicente and H. Pain. Motivation diagnosis in intelligent tutoring systems. In B. P. Goettl, H. M. Halff, C. L. Redfield, and V. J. Shute, editors, *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, pages 86–95, Berlin. Heidelberg, 1998. Springer.

[2] T. del Soldato. Motivation in tutoring systems. Tech. Rep. CSRP 303, School of Cognitive and Computing Sciences, The University of Sussex, UK, 1994.

[3] T. W. Malone and M. R. Lepper. Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow and M. J. Farr, editors, *Conative and Affective Process Analyses*, volume 3 of *Aptitude, Learning, and Instruction*, chapter 10, pages 223–253. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, 1987.

[4] J. M. Keller. Motivational design of instruction. In C. M. Reigeluth, editor, *Instructional-design theories and models: an overview of their current status*, pages 383–434. Lawrence Erlbaum Associates, 1983.

[5] C. Crowley. Tkreplay. version 1.1. Software downloadable at http://www.cs.unm.edu/~crowley, 1996. Available on-line on 18th November 1998.

[6] A. de Vicente and H. Pain. Eliciting motivation diagnosis knowledge. *In this volume*.

Towards 'motivating' instructional systems

Angel de Vicente*and Helen Pain
Division of Informatics, Artificial Intelligence
The University of Edinburgh
E-mail: [angelv,helen]@dai.ed.ac.uk

Recent findings show how important the role of emotions is in human intelligence. Therefore, not surprisingly, an increasing amount of research in artificial intelligence is paying attention to affective issues, and to how these can make artificial systems behave in a more 'human' way. Similarly, computer instruction systems that know how different affective states can influence their students' learning (and inversely, how particular practices in teaching can influence their students' affective states) could leverage their instructional quality by attempting to keep the students in a cognitively receptive state. In this paper, after reviewing related research in this area, we present the prototype affective instructional system we have developed together with some preliminary results. Finally, we introduce some of the research directions that we are currently pursuing.

Keywords: **Intelligent Tutoring Systems, Motivation detection, Motivational planning, Educational dialogue, Student modelling**

1 Introduction

Recent research is helping to discard the old idea that emotions and intelligence do not mix well. The influential research by Damasio [4], for instance, shows not only that emotions are not detrimental to intelligence, but rather, that we need emotions in order to be intelligent. This is creating a lot of interest in the area which has been called "Affective Computing", which is "computing that relates to, arises from, or deliberately influences emotions." [15].

This renewed interest on affective issues is also permeating the research area of computer instruction. We believe that this was long overdue, since as Goleman [9] puts it, "The extent to which emotional upsets can interfere with mental life is no news to teachers. Students who are anxious, angry, or depressed don't learn; people who are caught in these states do not take in information efficiently or deal with it well." (pp. 78).

*Supported by grant PG 94 30660835 from the 'Programa de Becas de Formaci ón de Profesorado y Personal Investigador en el Extranjero", Ministerio de Educaci ón y Cultura, SPAIN.

Thus, our main research interest lies in creating instructional systems that can engage and motivate students. While these concerns are becoming very popular, the approach taken in most cases involves using a particular technique or presentation approach which the developers of the system believe will keep the students engaged (see for instance [17]). In contrast, our approach follows the idea that "truly personalized instruction must be individualized along motivational as well as cognitive dimensions [...]" [12].

We intend to create instructional systems that (following human teachers' practice) are able to select the course of instruction according to not only the cognitive state of the student, but also her affective state.

Lepper and Chabay [12] also give some suggestions on what should be added to an instructional system so that it can emphathise with the student:

1. General social knowledge or "some general rules concerning the appropriateness of different sorts of social and motivational remarks in various situations" pp. 251.

2. Specific background knowledge about the individual student.

3. A component to offer choices to the student and to analyse his responses (for instance, if he would like help, an easier problem, etc.).

The main work to date which attempted to implement these ideas is that of del Soldato [8]. Her work was pioneering in this area and introduced the idea of adding a motivational modeller and a motivational planner to the traditional Intelligent Tutoring Systems (ITS) architecture. We are continuing this direction of research, investigating some of the unexplored areas and limitations of previous research.

As a framework to test different theories and ideas, we have developed a prototype instructional system which we describe in section 2. The following sections deal with the techniques to be used in a future implementation of the system. We finish with a discussion of our approach and some brief conclusions.

2 Description of MOODS Prototype System

2.1 Underlying concepts

We have based our prototype system on a flexible structure which can be seen in figure 1. This helps to avoid the development of generic motivating techniques that fail to deploy truly individualised instruction.

The structure can be seen as an extension to both the motivation diagnoser and motivational planner devised by del Soldato [8]. The instructional system can be seen as a system powered by a cycle in which the student is central in every step. The actions of the student (exercise outcomes, time used for solving a problem, etc.) are stored in an interaction history database. This database together with information about student's long term characteristics inform the motivation diagnoser, which updates the student's motivation model. Given the current motivation model of the

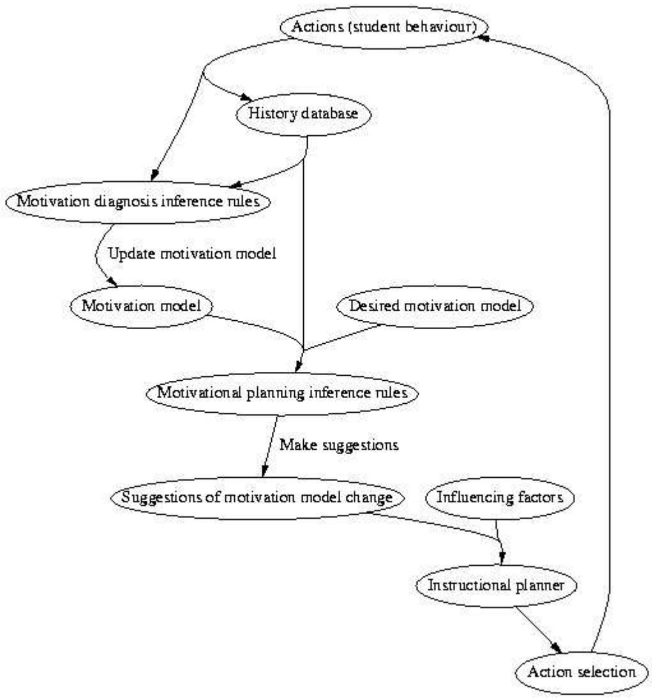


Figure 1: System structure

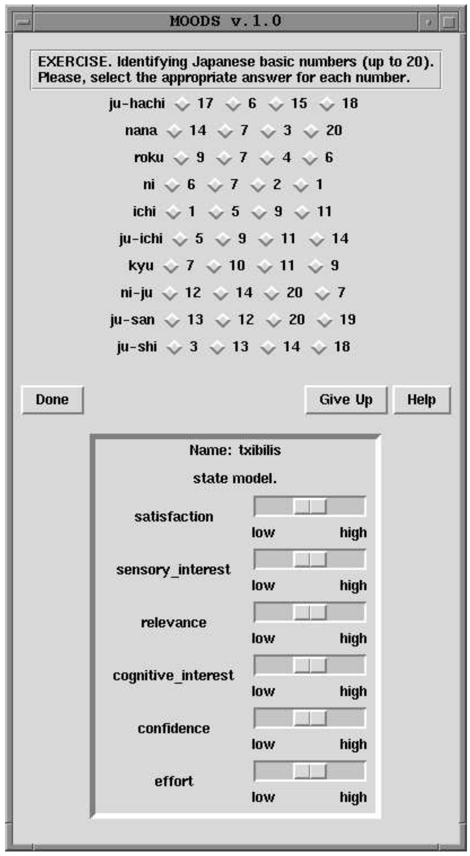


Figure 2: MOODS interface

student and the desired model, the motivational planner can make suggestions as to which motivation factors should be changed. Lastly, the instructional planner takes into account the suggested changes to the motivation model together with curriculum constraints information and information on student trait characteristics, in order to select the best available action to bring about the desired changes.

The structure results in a very modular approach, in which certain aspects can be left undeveloped without detriment to the correct functioning of the overall system, as we will see in later sections. The structure also allows for highly individualised instruction, which can modify itself to match each students' needs.

2.2 The MOODS prototype system

The structure described in section 2.1 has been partially implemented in a prototype instructional system, whose interface can be seen in figure 2. MOODS is a simple tutoring system to teach Japanese numbers up to 100. Currently the system consists of 10 lessons, ranging from simple presentation of numbers to memorise to a Tetris-like game to perform number additions[1]. Thanks to these different styles of lessons,

[1]To avoid making the task unnecessarily difficult, numbers were not taught using the actual Japanese sillabry. Instead, we used the *romaji* syllabry, which is a transcription into the latin alphabet of the Japanese

a number of instructional paths can be created, each of them consisting of a different combination of lessons.

As a central component of MOODS, we implemented a "motivation model" that can be seen in figure 3.

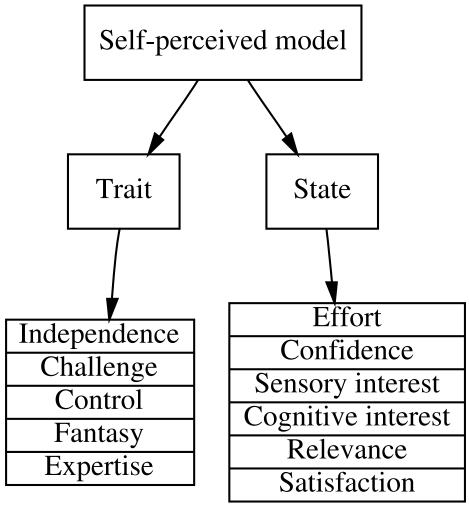


Figure 3: MOODS motivation model

The model is divided in two main categories: traits ('permanent' characteristics) and states ('transient' characteristics). This model is based primarily on two theories of motivation in education [13, 10] and on one of the very few motivational models implemented in an ITS to date [8].

Undoubtedly, there are many other factors from other accounts of motivation (e.g. [16, 18, 3]) and from previous motivation models implemented in ITSs (e.g. [14]) that could have been part of our motivation model. The task of choosing between these different factors is a difficult one, but we believe that our model presents a useful set of the main important characteristics for student motivation, while doing it with a small number of variables which makes it feasible to use.

Part of this model is present in the MOODS interface (figure 2) as will be explained in section 3.1. Otherwise, the system consists of very simple lessons in which the student has to memorise, fill in blanks, select answers (as in the example seen in figure 2), or play the Tetris-like game.

3 Implementation approaches in MOODS

In the next sections we present the approach taken for the main aspects of MOODS design: motivation diagnosis and motivational planning.

kana, and which is widely used to teach Japanese to westerners.

3.1 Motivation diagnosis

This is the area that so far we have paid more attention to, since previous research has been concerned mainly with motivational planning, leaving the detection of the student's motivational state virtually unexamined. The approach taken by del Soldato [8] used a number of rules to detect three student characteristics based on student's actions, but the rules seem to have been elicited through common sense, and their accuracy was not tested.

We intend to focus mainly on this effort, since it seems obvious that good detection must be necessary for good instructional planning. After reviewing possible motivation diagnosis techniques [5], we performed a study to check the viability of two of the techniques reviewed: questionnaires and self-report [7]. The questionnaire technique was used to gather information about a student's trait characteristics (see figure 3) prior to the interaction with the main MOODS interface, through a small on-line questionnaire.

The self-report approach was used to gather information about a student's motivational state throughout the interaction with MOODS via the manipulation by the student of six sliders, which represent the six 'state' categories of our motivation model. Each of these sliders (which can be seen in the bottom part of figure 2) has 5 different possible settings from 'low' to 'high'.

While the acceptance of these methods by the students was high, preliminary analysis of the data prompted some issues that will influence the future approach to motivation diagnosis [7]. Participants seemed to think that self-report could be a good method for communicating with the computer about motivational issues, and a method not intrusive on their learning. But also, as noted in the previous section, some of the students commented that a longer period of use of the system may make them lose interest in using the self-report facilities.

Given this, we think the best approach would be to reduce the number of self-report factors, concentrating on the ones that are easier to update and more regularly updated in our study. We believe that 'confidence, 'effort' and 'satisfaction' could probably be the factors that are given to the student for regular update, and use the less used factors for problem detection. Since 'sensory interest', 'cognitive interest' and 'relevance' are not used very often, these could be shown to the user only as a remedial action, when student's satisfaction is getting low. This modified approach would mean a mixture of self-report and expert system approaches as reviewed in [5]. We intend to test these suggestions in a future version of MOODS.

Thus, a modification of the actual system will mean that motivation diagnosis will be performed by a combination of four of the methods reviewed in [5]: questionnaires, verbal communication, self-report and expert system. Questionnaires will be used to gather data of long-term characteristics about the student. Self-report will be used to gather data about short-term characteristics that the students themselves are happy and confident of updating. A behind-the-scenes expert system will monitor the student's actions to infer certain motivational factors which are not easy for the student to update, and will try to correct any motivational problems (see section 3.2). The main means of communicating with the student will be verbal, and this is briefly discussed in section 3.3.

3.2 Taking remedial actions

Given the intrinsically difficult task of detecting a student's motivation without her collaboration [8] we attempt to use self-report methods to infer this. As seen in [7], this method is not totally appropriate either, since the student can fail to update the self-report facilities for various reasons (too difficult to decide own values; task too engaging; updating the sliders doesn't bring an obvious response from the system, which makes it apparently unhelpful, etc.). Therefore, we believe a combination of both is a more appropriate approach.

In our future version of MOODS, self-report facilities will be limited to the three factors which are more easily used by the students (i.e. confidence, effort and satisfaction). At the same time, the system will be implemented with a computer-generated motivation model of the student. This model will have the same factors as the student-updated model, plus the other factors which seem difficult for the student to update. The system will monitor both models (computer-generated and student-generated) for signs of motivation problems. Examples of situations that could prompt the system to actively ask the student about motivational problems could be: 1) if the values of the common factors in both models differ greatly (indicating perhaps a lack of interest by the student in updating the model, or a wrong motivation diagnosis by the system); 2) the values of the motivational factors are beyond an established threshold that indicates that remedial action is necessary.

Obviously, such a system needs to have knowledge about how to detect a student's motivation. In order to elicit appropriate diagnosis knowledge, we have planned a study that will help us to elicit this knowledge from experienced human teachers. In the 'teacher' mode, MOODS can be used to replay the actions of a previous student interaction[2]. By using this mode, we will let experienced teachers predict the student's motivational 'state' during the interaction. Interviews with the teachers while showing them a replay of their own predictions will help us to rationalise their diagnosis of the student's motivational state. The knowledge thus elicited, will be the basis for our modelling engine. More details about this study can be found in [6]

3.3 The role of language

The flexibility of language makes it a complicated and at the same time very powerful means of communicating affective information. Presently we are analysing educational dialogues in order to find how students and teachers communicate affective information.

Some examples of sentences which illustrate the sort of information that we would be interested in follow[3]. Each of these sentences carries a lot of affective

[2]The recording and replaying facilities were possible thanks to the TkReplay program [2].

[3]The dialogues from which these sentences are taken are part of the Vicarious Learner project [1], and some of them can be found at http://www.cogsci.ed.ac.uk/ rcox/drs_vicars_bak/drs_intro.html (on 27/2/98)

information, on which we can only comment briefly here due to lack of space:

1. TUTOR: . . . so if you'd like to start off by doing the first sentence. — *Suggestion of instruction. Challenge.*
2. TUTOR: Well, what word did you get rid of? — *Understanding hint*
3. TUTOR: That's looking good — *Encouragement*
4. STUDENT: . . . I'm not quite sure if I really know that or not from the rules . . . — *Confi dence on own abilities low*
5. STUDENT: Let me look at the rule for **a** — *Increase instructional plan control*

A natural language understanding and generation ability is outside the scope of our project, but we intend to use the main sentence categories (for example: hint, encouragement, etc.) found in the dialogue analyses as guides and templates to create sentences in natural language. Thus, the student will be offered an interface where she can select a sentence from a set, and the system will be able to act on this sentence accordingly (for instance, sentence 4 above would decrease the confidence variable in the internal motivation model).

Similarly, the sentence patterns will be used according to the motivational needs of the system at any given moment. For instance, if the remedial system (see section 3.2) detects that the confidence level is dropping too fast, the system could consider giving hints to the student, which in our case would consist of canned text for each lesson mixed with sentence patterns, that would result in sentences similar to number 2 above.

4 Conclusion

In this paper we have briefly presented the current status of our research. We introduced our developed prototype instructional system (i.e. MOODS), and we attempted to give a feeling for how the system will relate to students once a final version is developed.

For the moment, the main emphasis of our research has been on motivation diagnosis and this is reflected in this paper. There are still many issues to be solved, but we believe that our system and the framework presented allow us to implement each issue in a modular and flexible way.

Acknowledgements

We would like to thank Shari Trewin for useful comments on previous versions of this paper.

References

[1] Cox, R., McKendree, J., Tobin, R., Lee, J., and Mayes, T. (1998). Vicarious learning from dialogue and discourse: A controlled comparison. Under review for Instructional Science.

[2] Crowley, C. (1996). Tkreplay. version 1.1. Software downloadable at http://www.cs.unm.edu/˜crowley. Available on-line on 18th November 1998.

[3] Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*. Behavioral Science Series. Jossey-Bass Publishers, San Francisco.

[4] Damasio, A. R. (1996). *Descartes' error. Emotion, reason and the human brain*. Papermac, London. First published as a Grosset/Putnam Book 1994 by G.P. Putnams Sons, New York.

[5] de Vicente, A. and Pain, H. (1998). Motivation diagnosis in intelligent tutoring systems. In B. P. Goettl, H. M. Halff, C. L. Redfield, and V. J. Shute, editors, *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, volume 1452 of *Lecture Notes in Computer Science*, pages 86–95, Berlin. Heidelberg. Springer.

[6] de Vicente, A. and Pain, H. (1999a). Eliciting motivation diagnosis knowledge. In [11], pages 651–653.

[7] de Vicente, A. and Pain, H. (1999b). Motivation self-report in ITS. In [11], pages 648–650.

[8] del Soldato, T. (1994). Motivation in tutoring systems. Tech. Rep. CSRP 303, School of Cognitive and Computing Sciences, The University of Sussex, UK.

[9] Goleman, D. (1996). *Emotional intelligence: Why it can matter more than IQ*. Boomsbury, London. First published by Bantam Books, New York, 1995.

[10] Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth, editor, *Instructional-design theories and models: an overview of their current status*. Lawrence Erlbaum Associates.

[11] Lajoie, S. P. and Vivet, M., editors (1999). *Proceedings of the Ninth World Conference on Artificial Intelligence in Education*, Amsterdam. IOS Press.

[12] Lepper, M. R. and Chabay, R. W. (1988). Socializing the intelligent tutor: Bringing empathy to computer tutors. In H. Mandl and A. Lesgold, editors, *Learning Issues for Intelligent Tutoring Systems*, chapter 10, pages 242–257. Springer-Verlag.

[13] Malone, T. W. and Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow and M. J. Farr, editors, *Conative and Affective Process Analyses*, volume 3 of *Aptitude, Learning, and Instruction*, chapter 10, pages 223–253. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

[14] Matsubara, Y. and Nagamachi, M. (1996). Motivation system and human model for intelligent tutoring. In C. Frasson, G. Gauthier, and A. Lesgold, editors, *Proceedings of the Third International Conference on Intelligent Tutoring Systems*, pages 139–147, Berlin. Springer-Verlag.

[15] Picard, R. W. (1997). *Affective Computing*. The MIT Press, Cambridge, Massachusetts.

[16] Spitzer, D. R. (1995). *SuperMotivation*. AMACOM Books, New York.

[17] Stoney, S. and Oliver, R. (1998). Interactive multimedia for adult learners: Can learning be fun? *Journal of Interactive Learning Research*, **9**(1).

[18] Wlodkowski, R. J. (1998). *Enhancing adult motivation to learn*. Jossey-Bass, San Francisco.

A computational model of Affective Educational Dialogues

Angel de Vicente* and **Helen Pain**
Artificial Intelligence; Division of Informatics
Edinburgh University; Edinburgh (United Kingdom)
{angelv, helen}@dai.ed.ac.uk

Abstract

Issues of affect have been largely neglected in computational accounts of dialogues so far. Given the importance of affect in education, we present in this paper on-going work on a computational model of affective educational dialogues.

Introduction

Emotional and affective issues are recently enjoying an increased amount of attention in the field of Artificial Intelligence, as can be seen by the number of workshops and conferences devoted to them (e.g.: (Emotion in HCI 1999; Emotion-based Agent Architectures 1999; Affect in Interactions 1999)). These issues are also of great importance in the area of Educational Dialogues, both for better understanding human educational interactions and for creating Tutoring Systems that communicate with students in a more natural and effective way.

Language is a powerful communicator of affect, and it can be used in tutoring systems to empathise with students and to detect their emotional state during the interaction. Related work exists (e.g. (Allport 1992; de Rosis & Grasso 1999; Horvitz & Paek 1999; Person *et al.* 1999)), but a detailed computational model of affective dialogue is missing, a gap that we attempt to fill with the on-going work reported in this paper. The main aim of the model presented in this paper is to generate educational dialogues, focusing primarily on their affective characteristics.

The model (discussed in more detail below) bases its decisions on a number of sources: rules drawn from several theories of motivation and education; features of the past interaction of the student and a model of the student himself. In its current version it has been incorporated into a mock instructional system called AFDI (AFfective DIaloguer), described below. We also give and comment on a short example of a dialogue generated with the current model.

*Supported by grant PG 94 30660835 from the "Programa de Becas de Formación de Profesorado y Personal Investigador en el Extranjero", Ministerio de Educación y Cultura, SPAIN.

A computational model of affective educational dialogues

One of the difficulties of creating a computational model of affective educational dialogues is that it cannot be created in isolation from an instructional system. Although the system does not have to be a 'real' tutoring system, the model needs information about an instructional interaction in order to be useful. In this sense, the development of an educational dialogue model amounts to the development of an instructional model in which different parts of it are 'glued' together by the use of language. This can be seen in figure 1, which represents a simplified overall view of the AFfective DIaloguer (AFDI) computational model.

The rectangular nodes represent steps of text generation; the oval nodes represent decision-making points and the skewed rectangular nodes represent student interaction. Thus, the model assumes that an instructional interaction will start with a general introduction, followed by a cycle in which different instructional units are studied, followed by some concluding remarks. During the main instructional cycle, the model allows five different types of interventions: two of them initiated by the system (to provide help or to attempt to solve motivational problems); and the other three initiated by the student (to finish the lesson, to give up or to ask for help).

The oval nodes in the model represent, as mentioned, decision points. It is in these nodes that affective issues are taken into account in order to generate the dialogue. These decisions are based on the history of interaction with the system and the following student characteristics[1]:

- Student's trait characteristics: fantasy, challenge, control, independence. The model uses a measurement of how much the student appreciates the given characteristic during instruction. For example, does the student like challenging situations?
- Student's motivational state characteristics: satisfaction, relevance, confidence, effort, sensory interest, cognitive interest. These characteristics are transient, and the model assumes two different sources for them:

[1]These are based on previous work on diagnosis of student's motivation (de Vicente & Pain 1999)).

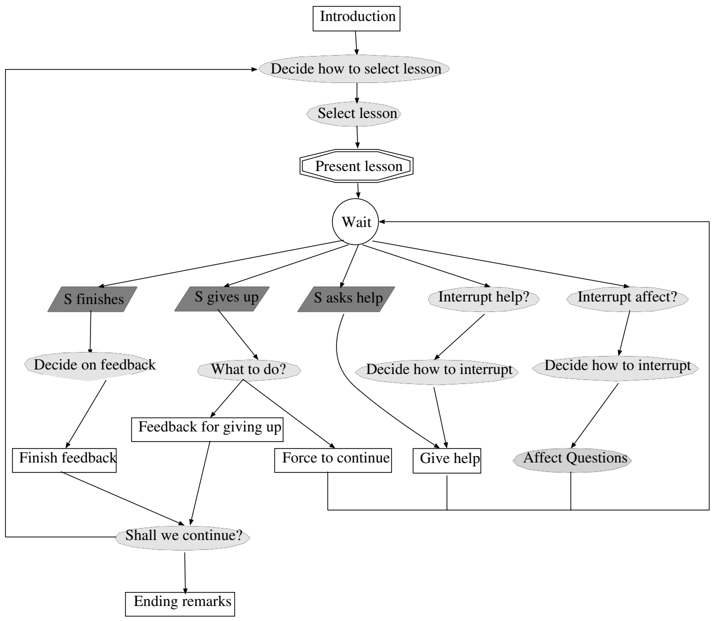


Figure 1: AFfective DIaloguer (AFDI) Computational model (simplified)

– Self-report, in which the student provides his own estimation of each of them.
– Inferred from the behaviour of the student.

Language used

In order to keep complexity low, our generation process is oriented towards generating affective educational dialogues, but producing the actual text from canned text options. A more sophisticated method of text generation would be desirable, but our main concern in developing this model is to present a plausible affective dialogue generation approach in the field of instructional systems[2]. Nevertheless, the design of the model (and of the AFDI system) is modular, and it would allow for an inclusion of a more sophisticated Natural Language Generation engine.

The tutor dialogue moves are selected from a set of possible sentences which are classified according to their content and their affective characteristics. The basic possible themes to generate tutor moves are given by the rectangular boxes in figure 1 (e.g. "Introduction", "Give help", etc.). For each tutor move, a number of possible student replies are attached. The selection of which particular tutor move to generate, or which possible replies to present to the student, is determined by the particular characteristics of the student and the history of the interaction.

When the model is in a text generation step (e.g. 'Introduction'), the program issues a request to generate a tutor move of certain characteristics (i.e. topic: help; subtopic: lesson1; flow: provide; politeness: high,; etc.)[3]. Another procedure is responsible for returning an appropriate move of those characteristics. In the current AFDI version this is done by looking at a database of "typical" tutor moves.

Tutor moves Given a request for a particular type of tutor move, our model will generate a move that matches the given requirements among all the possible tutor moves. Some examples of possible tutor moves are given in table 1(a). The three columns of this table represent respectively: the move characteristics; the actual text, and the student reply options.

In order to add more flexibility to the language used in the model, the text can be pre-determined or selected during run-time. Thus, in the first tutor move in table 1(a), the text is made of a variable part ([sel GEN_INTRO]), which selects randomly from a list of possibilities for generic introductions. These (and examples of other sub-sentence variables) are given in table 1(b). This allows for a more varied dialogue output, while keeping the moves database simple.

Student replies The tutor moves also have information about which student replies are appropriate for each of them. For example, in the first move in table 1(a), this is given by

[2] For an approach in which Natural Language Generation techniques are used for affective text generation see (de Rosis, Grasso, & Berry 1999; de Rosis & Grasso 1999) and (Porayska-Pomsta, Mellish, & Pain to appear 2000) for more linguistically motivated research on generating teachers' language in educational dialogues.

[3] The politeness characteristic is not used in the current version of AFDI.

	Characteristics	Text	Student reply
1	introduction:intro	[sel GEN_INTRO]	reply_intro
2	feedback:pos_inc_conf	[sel POS_FEED] [sel INC_CONF]	reply_pos_inc_conf
3	help:talk_about_help	It seems that perhaps you would benefit from some help. Would you like some?	reply_y_n
4	ending:due_system	Well, I think is perhaps time to leave it for today.	reply_bye

(a) Sample tutor moves

GEN_INTRO	Welcome to Moods, your affective tutor! Welcome to Moods Good morning, and welcome to Moods
INC_CONF	You see, you can do it! And you thought you could not do it?
POS_FEED	Well done! Congratualations, you are doing great.

(b) Sample sub-sentences

reply_intro	Hi Thanks
reply_y_n	Yes No
reply_pos_inc_conf	Thanks, it is true. It is easier than I thought! Well, it was probably just luck!

(c) Sample student replies

Table 1: AFDI language

the variable 'reply_intro'. This variable represents the possible student replies, which can be generated at run-time or pre-determined. In the current AFDI version, only the selection of lessons by the student is generated at run-time (based on the previous history of interaction), while the other possible replies are predetermined. Examples of possible student replies are given in table 1(c).

The affective knowledge

The main aspect of AFDI is its ability to make affect related decisions in order to generate the dialogue. These decisions are guided by the affective knowledge, which is implemented by two types of knowledge-based rules: *dialogue planning rules* (those shown in figure 1 as elliptic nodes) and *modelling rules*.

The *dialogue planning rules* represent knowledge about the generation of Educational Dialogues moves, given the history of interaction and the student's characteristics. The *modelling rules* represent knowledge about which information concerning the student's affective state can be inferred from his interaction with the system. These two types of rules are discussed in the following sections.

Dialogue planning rules The dialogue planning rules implemented in the current version of AFDI are very simple. For brevity, we give here just one example in detail and then present a summarised version of all the rules in the current AFDI version.

The rule *stud_choose* shown in table 2 represents a dialogue planning rule concerned with *control*, a major affective factor during an instructional interaction. In AFDI the selection of the next lesson to study can be done entirely by the student, entirely by the system, or by the collaboration of both. This is reflected in the rule *'stud-choose'*. There we

control::trait	**relevance**::state	Output
very high	any value	stu_chooses_all
average–high	below very high	stu_chooses_type
below average	below average	stu_chooses_type
average–high	very high	system_chooses
below average	average or higher	system_chooses

Table 2: Rule stud_choose

can see that the decision at this point is influenced by two variables: the student's *control* trait characteristic, and the *relevance* state characteristic. Thus, we see that if the student's *control* is very high[4], then we always let the student choose the next lesson. If his *control* is *average* or *high*, then the decision depends on whether he thinks that the instruction is of relevance to him. If it is *very relevant*, then we assume the system is doing a good selection of lessons, and we should let the system choose the next lesson to perform. If the student thinks that the materials are not really relevant to his learning, then we let him choose the type of exercise to do next.

A summarised version of all the rules in the current AFDI version are given in table 3.

Modelling rules These rules are similar in syntax to the dialogue planning rules but they are formalisations of the inferences about the student affective state that can be drawn from the student's replies. As explained earlier, every time that the system generates some text, the student is offered

[4]The values for all the variables in the student model can take five different values: -10, -5, 0, 5 and 10, corresponding respectively to: very low, low, average, high, very high.

Rule name	Depends on	Description
stud_choose	control, relevance	Who chooses next lesson (student, partially student, system)
decide_feedback	performance, effort, confidence	What type of feedback to give
student_chooses_to_continue?	how_many_lessons, control	Should the student decide whether to continue?
system_continue_instruction?	how_many_lessons, decaying_effort	Should we continue the instruction?
allow_give_up?	effort, challenge	Should we allow the student to give up?
interrupt_affect_R1	satisfaction	Is general satisfaction so low that we should interrupt the student?
interrupt_affect_R2	sensory_interest, cognitive_interest	Is the interest so low that we should interrupt the student?
interrupt_affect_R3	relevance	Is the material being taught irrelevant to the student's needs?
interrupt_affect_R4	Inferred user_state, Reported user_state	Is the reported model very different from the inferred one?
student_decides_interrupt_affect?	independence	Should the student be asked about whether to solve a motivational problem?
interrupt_help_R1	confidence	Is the confidence so low that we should interrupt to offer help?
student_decides_interrupt_help?	independence	Should the student be asked whether he wants the help or not?

Table 3: Dialogue planning rules

some choices to form his reply. According to this reply, we can sometimes infer certain affective information that can be used to update his affective model. These rules encapsulate some of this knowledge. As for the dialogue planning rules, we present first one of them in detail.

STUDENT REPLY	CHANGES
easy	effort -5
hard	effort 5

Table 4: Modelling rule pos_plain

The rule *pos_plain* in table 4 is a formalisation of the inferred changes in the student model after the system has provided *pos_plain* feedback (positive feedback). The system can provide positive feedback with one of four possible moves (for instance "That was very good!"). To this, the student can reply with two different replies:

1. Thanks, it was easy.
2. Thanks, but it was hard!

Thus, based on the student's reply, this rule helps to infer the amount of effort that the student has put into the task. A "Thanks, it was easy" reply means to our model that the student did not put too much effort into the task[5], while a reply of "Thanks, but it was hard" would mean that the student made a big effort[6].

AFDI has other similar rules to update the student model after replies to various types of feedback, help, etc., but a description of all of them would take too much space in here. We illustrate some of them below with an example of a dialogue generated with AFDI.

Implementation of the model

In order to experiment with the model, we have implemented it in a simple application, whose interface can be seen in figure 2. The working of the system is very simple. In the top part of the interface the representations of the student's trait and state (both inferred and reported) characteristics are given. By having them available all the time we can modify them easily to see their effect on the generation of the dialogue. Similarly, we can also see how different student replies affect the student model. In the top part of the interface, a brief description of the simulated lesson is also given.

The bottom part contains two frames: 1) *Dialogue Log*, where all the dialogue generated and the rules applied to generate it are logged; 2) *Dialogue move*, where the current tutor move and its possible student replies are displayed. In debugging mode (as seen in figure 2) there is a third frame where the history of interaction database is shown as the simulated instruction takes place. In this database we store information about the lessons studied, the outcome for each of them, etc.

In order to interact with the system, we simulate the behaviour of a student[7] and decide on the possible outcome to each lesson (i.e. succeed, give_up, etc.). The time that the student would spend in studying the lesson and a measurement of performance is also simulated through the se-

[5] We decrease the student model's value of effort by 5.

[6] We increase the value of effort by 5.

[7] Using an approach which we would rather call 'Dorothy' than 'Wizard of Oz' ...

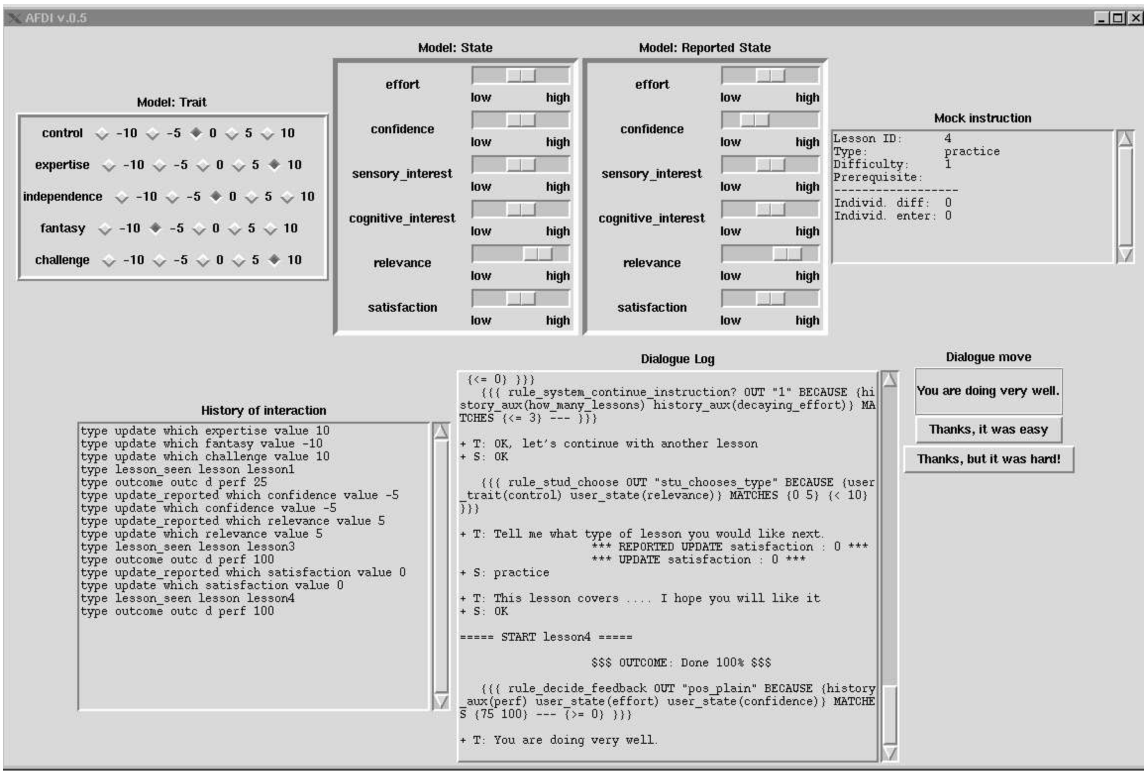


Figure 2: Main interface of AFDI.

lection of the appropriate menu options. As the interaction takes place, the student model will vary depending on the student's performance, the replies to the tutor moves, etc. These changes will also be reflected in the selection of the lessons and future dialogue moves. This enables us to see the model functioning in an interactive mode. An example dialogue generated in this way is given and commented upon in the following section.

AFDI Generated Example Dialogue

Below we present an abridged version of an actual dialogue generated with AFDI. For conciseness, we have omitted certain parts of the dialogue and present only the sections that indicate more clearly the main aspects of AFDI. The text is divided in two columns: to the left, the actual dialogue is presented; to the right, we present a summary of the interaction with the system plus a description of the *dialogue planning* and *modelling* rules that were used to shape the dialogue.

After a brief introduction and the update of the trait characteristics by the (simulated) student, the first point of interest is that marked in the dialogue as $\boxed{1}$. This illustrates how the important motivational factor of *control* (or, more importantly, *feeling of control*) is dealt with in AFDI. The rule used here was described in detail above. At this point AFDI decides that the student should be given some, but not total, control over the next lesson to study. This is motivated by the fact that the student's desire for control is not very high[8], and that the materials were not very relevant to him[9]. Thus, the student is asked to choose the difficulty of the next lesson to study.

The option of exercising *control* over the interaction is also seen in the dialogue at point $\boxed{11}$. There we see that despite reaching the established maximum number of lessons for this session (4 in this example), the student is offered the choice to decide whether he would like to continue with the instruction, as his *control* is high (greater than 0).

Control is also the concern in point $\boxed{5}$, but in this case AFDI limits the *control* exercised by the student. Because the student's desire for control is not very high, he has not put much effort into the task, and he likes challenging situations (challenge greater than 0), the system tries to encour-

[8] `user_trait(control) {0 5}` meaning that this variable had one of these values at this point.

[9] At the beginning of the interaction, the variables are initially set to *average* (or 0).

age him to try a bit harder.

Another issue which is of crucial motivational importance is feedback. Whether the feedback is positive or negative, whether it tries to encourage the student, etc. can affect how the student feels about the instruction. An example of this is given in points $\boxed{7}$ and $\boxed{9}$. In $\boxed{9}$ the feedback is positive, since the student's performance was very high (75% or higher; 100% in this example). But in $\boxed{7}$ we see that, despite similar performance, student's self-confidence is low (less than 0), and thus the feedback is characterised as 'pos-inc-conf' (positive feedback, plus increase confidence). This is realized in our example as "Well done! And you thought you could not do it?"

As seen above, the purpose of the feedback 'pos-inc-conf' is to increase student's self-confidence, but it is equally important to understand how the student reacts to it. Given the example tutor move in previous paragraph, the student can select from two possible replies: 1) "Thanks, it is true. It is easier than I thought!"; 2) "Well, it was probably just luck!". These are meant to provide information about whether or not the intended tactic had the desired effect. In this case, the student has selected the first reply, which would indicate (as given by the rule 'pos-inc-conf' in $\boxed{8}$) that his confidence has increased slightly (by a step of 5 in this example). If the other option had been chosen (i.e. "Well, it was probably just luck!"), the model would infer that the confidence was actually decreasing. Other points in the dialogue where other *modelling* rules are fired are $\boxed{3}$, $\boxed{4}$, $\boxed{6}$ and $\boxed{10}$.

In $\boxed{2}$ we illustrate the 'Interrupt' options of our model. In (de Vicente & Pain 1998) we have argued that in order to tackle motivational problems, an affective tutor would benefit from an ability to interrupt the instruction if the conditions required it. This ability to interrupt the instructional interaction, but not following a predetermined path, has two advantages:

1. It can create the illusion of a more flexible tutor who does not follow a strict instructional plan.

2. It can detect motivational problems as soon as they occur and it can take remedial actions.

AFDI checks regularly[10] whether it should interrupt. We see in $\boxed{2}$ one of these situations. The student has updated the reported variables of *sensory interest* and *cognitive interest* to *very low*. This indicates that the student does not find the instruction interesting at all, and therefore the rule *interrupt_affect_R2* is fired, which starts a sub-dialogue to find out the type of instruction that the student would prefer[11]. Thanks to this capability of interrupting, the system can offer help and can try to solve motivational problems as soon as their need is detected. At the same time, it would encourage the use of the self-report facilities, since the student would perceive that his interaction with the interface have immediate effect (de Vicente & Pain 1999).

[10] Every 10 seconds in the current version.

[11] This is preceded by a decision to ask the student whether he would like this interruption, since his 'control' is high.

Conclusions

In this paper we have presented a computational model of educational affective dialogues, which is being developed in the context of developing 'Affective Tutors'. This model does not attempt to present a psychologically plausible model of affective communication, but rather to give a first approximation of the neglected issue of affect in Educational Dialogues.

Currently the model generates simple, although plausible, educational dialogues. Nevertheless, evaluation of them is necessary. We are planning an evaluation of the model in which excerpts of dialogues will be presented to tutors, in order to evaluate the appropriateness of the tactics used in our model. We hope that this evaluation will help to refine the model and to create a more appropriate model of affective Educational Dialogues.

Acknowledgements

We would like to thank the anonymous reviewers and our colleagues in the AI and Education Group at the University of Edinburgh for useful comments on previous versions of this paper.

Raw dialogue text (abridged)

Tutor: Welcome to Moods, your affective tutor!
Student: Hi
Tutor: Please, let me know about you by filling in the questionnaire on personal traits.

```
UPDATE control : 5
UPDATE expertise : -5
[...]
```

Student: OK, it is done
Tutor: Shall we start? The material to be learnt covers basic Japanese numbers.
Student: OK

```
rule_stud_choose
user_trait(control) {0 5}
user_state(relevance) {< 10}
---> stu_chooses_type
1
```

Tutor: Tell me what type of lesson you would like next.
Student: difficult

```
CALCULATING AVAILABLE LESSONS
```

Tutor: Let's now study .... Good luck
Student: OK

```
REPORTED UPDATE sensory_interest : -10
UPDATE sensory_interest : -10
REPORTED UPDATE cognitive_interest : -10
UPDATE cognitive_interest : -10
```

```
rule_interrupt_affect_R2
user_state(sensory_interest) {< 0}
user_state(cognitive_interest) {< 0}
---> 1
2
```

```
rule_student_decides_interrupt_affect?
user_trait(control) {> 0}
---> 1
```

Tutor: I'm a bit worried about your motivation. Would you like to talk about this?
Student: Yes
Tutor: Apparently you are not enjoying this class very much. Is this correct?
Student: Yes
Tutor: I see ... Maybe I could do better if you answer the following questions.
Student: OK
Tutor: You would like more ...
Student: theory?

```
affect_presentation
STUDENT_REPLY MATCHES theory
---> fantasy -5
3
```

Tutor: Thanks, I will try to follow your advice. Now, would you like to continue with this lesson?
Student: No

```
[...]
```

Tutor: OK, let's continue with another lesson
Student: OK
[...]
Tutor: In this lesson we will study .... Good luck
Student: OK

```
OUTCOME: Help (slow) 25%
```

Tutor: Look at .... Perhaps that helps you
Student: I already knew that, but thanks

```
help
STUDENT_REPLY MATCHES already
---> relevance -5
[4]
```

```
OUTCOME: Give up 50%
rule_allow_give_up?
user_trait(control) {< 10}
user_state(effort) {< 10}
user_trait(challenge) {> 0}
---> 0
[5]
```

Tutor: Come on, you cannot give up now. You have to try a bit harder
Student: OK, but I'm not sure I can do it

```
force_continue
STUDENT_REPLY MATCHES sure
---> confidence -5
[6]
```

```
OUTCOME: Done (slow) 75%
rule_decide_feedback
history_aux(perf) {75 100}
user_state(effort) ---
user_state(confidence) {< 0}
---> pos_inc_conf
[7]
```

Tutor: Well done! And you thought you could not do it?
Student: Thanks, it is true. It is easier than I thought!

```
pos_inc_conf
STUDENT_REPLY MATCHES easier
---> confidence 5
[8]
```

Tutor: OK, let's continue with another lesson
Student: OK
[...]
Tutor: Let's now study .... I hope you will like it
Student: OK

```
OUTCOME: Done 100%
rule_decide_feedback
history_aux(perf) {75 100}
user_state(effort) ---
user_state(confidence) {>= 0}
---> pos_plain
[9]
```

Tutor: That was very good!
Student: Thanks, it was easy

```
pos_plain
STUDENT_REPLY MATCHES easy
---> effort -5
[10]
```

```
[...]
rule_student_chooses_to_continue?
history_aux(how_many_lessons) {> 3}
user_trait(control) {> 0}
---> 1
[11]
```

Tutor: Perhaps you would like to finish now?
Student: Yes
Tutor: I hope you learnt something useful. See you next time.
Student: Bye

References

Affect in Interactions 1999. International Workshop on Affect in Interactions. Towards a New Generation of Interfaces (Annual Conference of the EC I3 Programme). Available at the World Wide Web: http://gaiva.inesc.pt/i3ws/i3workshop.html (October 18, 1999).

Allport, G. 1992. Towards a computational theory of dialogues. CSRP Report 92-3, University of Birmingham. School of Computer Science. Cognitive Science Research Centre.

de Rosis, F. and Grasso, F. 1999. Affective text generation. In Affect in Interactions (1999). Available at the World Wide Web: http://gaiva.inesc.pt/i3ws/accepted-papers.html (November 25, 1999).

de Rosis, F., Grasso, F. and Berry, D. C. 1999. Refining instructional text generation after evaluation. *Artificial Intelligence in Medicine* 17(1):1–36.

de Vicente, A. and Pain, H. 1998. Motivation diagnosis in intelligent tutoring systems. In Goettl, B. P., Halff, H. M., Redfield, C. L. and Shute, V. J., eds., *Proceedings of the Fourth International Conference on Intelligent Tutoring Systems*, volume 1452 of *Lecture Notes in Computer Science*, 86–95. Berlin. Heidelberg: Springer.

de Vicente, A. and Pain, H. 1999. Motivation self-report in ITS. In Lajoie, S. P. and Vivet, M., eds., *Proceedings of the Ninth World Conference on Artificial Intelligence in Education*, 648–650. Amsterdam: IOS Press.

Emotion-based Agent Architectures 1999. Workshop on emotion-based agent architectures. Available at the World Wide Web: http://www.ai.mit.edu/people/jvelas/ebaa.html (October 18, 1999).

Emotion in HCI 1999. A British HCI Group one-day meeting in conjunction with University College London on “Affective Computing: The Role of Emotion in HCI”. Available at the World Wide Web: http://www-users.york.ac.uk/~am1/affective.html (April 8, 1999).

Horvitz, E. and Paek, T. 1999. A computational architecture for conversation. In Kay, J., ed., *Proceedings of the Seventh International Conference on User Modeling*, 201–210. Wien New York: Springer. Available at the World Wide Web: http://www.cs.usask.ca/UM99/ (May 24, 2000).

Person, N., Klettke, B., Link, K. and Kreuz, R. 1999. The integration of affective responses into AutoTutor. Available at the World Wide Web: http://gaiva.inesc.pt/i3ws/accepted-papers.html (June 5, 2000).

Porayska-Pomsta, K., Mellish, C. and Pain, H. to appear, 2000. Aspects of speech act categorisation: Towards generating teachers’ language. *International Journal of Artificial Intelligence in Education. Special Issue on Analysing Educational Dialogue Interaction* 11.

Informing the Detection of the Students' Motivational State: an Empirical Study

Angel de Vicente and Helen Pain

Institute for Communicating and Collaborative Systems; Division of Informatics,
The University of Edinburgh
{angelv,helen}@dai.ed.ac.uk

Abstract. The ability to detect the students' motivational state during an instructional interaction can bring many benefits to the performance of an Intelligent Tutoring System (ITS). In this paper we present an empirical study which provided us with a considerable amount of knowledge regarding motivation diagnosis. We show how this knowledge was formalised in order to create a set of motivation diagnosis rules that can be incorporated into a prototype tutoring system. We also briefly present how these motivation diagnosis rules were evaluated in another study.

1 Introduction

Many tutoring systems attempt to motivate the student by using multimedia, games, etc. This approach seems to be based on the idea that it is possible to create instruction that is motivating *per se*. However, as Keller mentions in [6], it is not always the case that if the instruction is of good quality motivation will follow. There has been some previous work that has dealt explicitly with motivation in ITSs (e.g.[4, 5, 9]), but these have dealt mainly with instructional planning, and not so much with motivation diagnosis. We believe that there is a pressing need for more research in this area [2], and we focus in this paper on how to elicit and formalise knowledge to diagnose the student's motivational state during an interaction with an instructional system.

The issue of how human teachers detect their students' motivation has been virtually unexplored in AI and Education research. Introspection and observational studies could throw some light into this issue, but they may be of limited usefulness for motivation diagnosis in ITSs. In a traditional instructional setting or any other social interaction there is a vast amount of information available through various communication channels, such as facial cues, posture, etc. [1], and there has been some interesting research on incorporating some of these cues in instructional systems (e.g. [10]). But many such cues that help us detect other people's emotions (or motivation) are perceived unconsciously, which makes it difficult to elicit emotion (or motivation) detection knowledge.

In order to limit the amount of sources of information available for knowledge elicitation, we designed a study in which a tutor is asked to infer a student's motivational state. But in order to do so, the only information that is available to her[1] is the prerecorded screen interaction of a student with an instructional system. That is, the tutor

[1] In this paper there are two main characters: the participants of the study, and the students whose interactions with an instructional system were replayed. In order to facilitate readability we use

is only able to see on a computer screen the interface of the instructional system which the student was manipulating. We expected that it would be easier for tutors to rationalise their motivation diagnosis knowledge in this setting than if they were presented, say, with video-recordings of tutoring interactions. At the same time, we believed that the knowledge inferred in this way would be easier to formalise in terms of information available to the instructional system (such as the duration of the interaction with the system, mouse movements, etc.).

2 Background on recorded interactions

As explained above, the participants of this study watched a number of recorded interactions of a student with a prototype ITS for learning Japanese numbers. This prototype ITS (MOODS) is a simple tutoring system with an added motivation self-report facility. This means that the student is able to report about his perceived motivational state during the interaction with MOODS. As a result of a previous study [3], we had a number of recorded interactions with the system, which were used as the basis for the study presented in this paper.

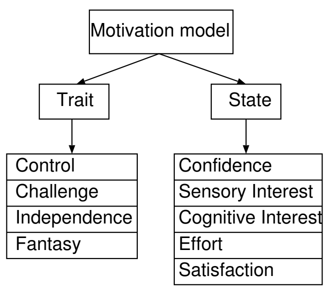


Fig. 1. Motivation model

The self-report facility available in MOODS is based on a motivation model which is presented in figure 1[2]. The model is based on the relevant literature (e.g. [6–8]) and it is composed of a number of motivational factors, which are divided in two classes: trait variables, or 'permanent' characteristics of the student; and state variables, or more 'transient' characteristics. Definitions for all these variables are given in table 1.

The trait variables in our motivation model are: control, challenge, independence and fantasy. There seems to be agreement on the importance of control and challenge for student's motivation. Fantasy, although not often included in theories of motivation

female pronouns to refer to the study participant and male pronouns to refer to the students. This has no relation to their actual gender.

[2] A more detailed description can be found in [3].

Variable	**Definition**
Control	Refers to the degree of control that the student likes having over the learning situation (i.e. does he like to select which exercises to do, in which order, etc. rather than let the instructor take these decisions?).
Challenge	Refers to the degree that the student enjoys having challenging situations during the instruction (i.e. does he like to try difficult exercises that represent a challenge for him?).
Independence	Refers to the degree that the student prefers to work independently, without asking others for help (i.e. does he prefer to work on his own, even if he finds some difficulties, and try to solve them by himself rather than asking for collaboration or help from others?).
Fantasy	Refers to the degree that the student appreciates environments that evoke mental images of physical or social situations not actually present (i.e. does he like the learning materials being embedded in an imaginary context?).
Confidence	Refers to the student's belief in being able to perform the task at hand correctly.
Sensory interest	Refers to the amount of curiosity aroused through the interface presentation (i.e. appeal of graphics, sounds, etc.).
Cognitive interest	Refers to curiosity aroused through the cognitive or epistemic characteristics of the task (i.e. regardless of the presentation issues, does the student find the task at hand cognitively appealing?).
Effort	Refers to the degree that the student is exerting himself in order to perform the learning activities.
Satisfaction	Refers to the overall feeling of goal accomplishment (i.e. does the student think that the instruction is satisfying and that it is getting him closer to his goals?).

Table 1. Definitions of motivation model variables

in Education, seems to be a factor that can play an important role in engaging the student (e.g. [8]). Independence, as defined in table 1, is related to challenge, but also to interpersonal motivations, such as: cooperation, competition and recognition [8].

The state variables represent transient characteristics of the student that relate to the material being learned. In figure 1 the state variables are presented in a more or less 'chronological' order. Thus, considerations of how confident he feels about succeeding in the task will likely take place before engaging in the task. This, together with the interest (both sensory and cognitive) that the lesson arouses in him, will influence the effort that he will put into the task. Satisfaction, as defined in table 1, represents the overall feeling of goal accomplishment, and will be influenced by all the variables above, plus by the outcomes of the task [6].

3 Materials

The participants of this study were asked to watch the recorded interactions of a student with MOODS, and they were asked to infer and comment on the motivational state of the student during the interaction. For this, we developed A_MOODS, which can be used to replay the actions of a previous student interaction with MOODS and to predict his motivational state. The A_MOODS interface can be seen in figure 2.

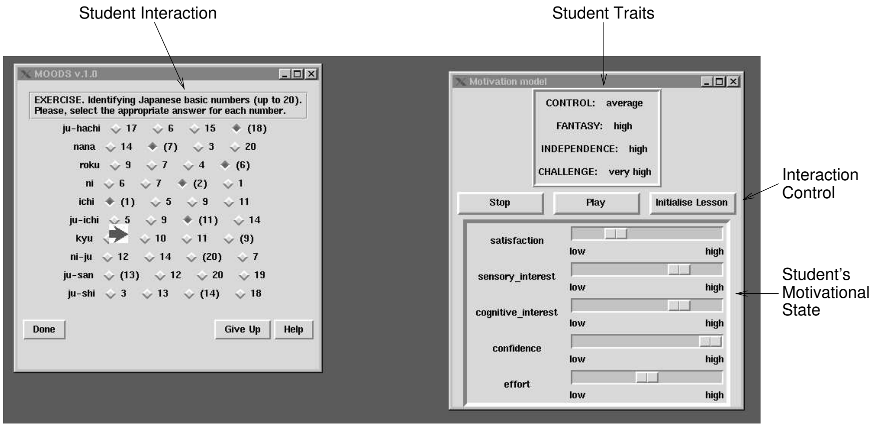


Fig. 2. A_MOODS interface.

The actions made by the student are replayed in the window to the left (with title MOODS v.1.0). To facilitate the viewing of the student interaction, an arrow (around the centre of the window in figure 2) indicates the mouse movements.

The window to the right (titled Motivation model) consists of three frames.

1. The top frame is a representation of the *student traits* of the motivation model (described in section 2). The values for these were obtained through a questionnaire during the self-report study described in [3].
2. The three buttons in the middle of the window control how the student interaction is replayed.
   – The replay of the student interaction starts when *Play* is pressed, and stops when *Stop* is pressed.
   – The button *Initialise Lesson* allows the user to 'rewind' the interaction to the beginning of the current lesson.
   – The replay of the interaction is done in real time, except for the replay of the theory lessons. In this type of lesson there is very little student activity, and therefore the system simply shows a message regarding how long the student took to learn the lesson[3]. After this the interaction continues.
3. The bottom frame is a representation of the student's *motivational state*, as discussed in section 2 (The task of the participants for this study was to predict the likely values of these motivational variables as the instruction replay took place).

[3] A sample message of this type is: "The student used 147 seconds to study this lesson".

In order to let the participants update the student's motivational model, the interaction stops by default in any of the following three situations:

1. When the student presses any of the buttons (Done, Give Up, or Help), but before any feedback is given by the system.
2. After feedback is presented to the student.
3. When a new lesson is presented to the student.

In these cases a small message[4] is shown to remind the participant to update the motivational model. If the participant does want to update the motivational model or comment on any aspect of the instruction at any other time, she can stop the interaction by pressing the *Stop* button.

4 Methodology

For this study 10 post-graduate students with previous teaching and/or tutoring experience volunteered to participate. After reading the study instructions, the interaction with A_MOODS started. The interaction can be summarised as follows:

1. The participant was given information about the trait characteristics of a student.
2. Then she was shown a replay of the student's interaction with MOODS.
3. Throughout the interaction, and particularly at any stop points, the participant was encouraged to give verbal comments on the student's motivational state and the possible factors affecting it.
4. Whenever the interaction was paused the participant was asked to update the motivational state if she had enough information to make an inference. And she was asked to verbalise the reasoning behind her inferences.
5. When the student pressed any of the three buttons available (Done, Give up, or Help), the participant was also encouraged to comment on the type of feedback that she thought would be the most appropriate to give to the student at that moment.

Throughout the duration of the study, the participant's comments were recorded on an audio disc, and were later transcribed and analysed.

5 Results

Before the study started, most of the participants commented on the perceived difficulty of the task. They expected not to be able to make any inferences based on the information provided. Contrary to their expectations, most of the participants made a considerable number of reasoned inferences about the student's affective state, and at the end of the study they commented that the task was actually not so difficult and that there was quite a lot of information available to them in order to perform these inferences.

On average, the time participants spent with the system was around 36 minutes. In this time, an average of 4 lessons were covered, and 8.5 inferences were made per participant. In total we collected 85 inferences. The excerpt below illustrates the type of comments that participants made during this study.

[4] "Please update the motivational model. Afterwards press 'Play' to continue."

```
Interviewer:
```

```
And [...] why do you think he is satisfied at this point?
```

```
Participant:
```

```
Well, [...] he is hovering the mouse over the answers each time, he wasn't
randomly moving the mouse, he is looking for the answer, [...] and that he
didn't take a long time to answer the questions. To me that would suggest
that the task is interesting enough to complete with some attention and to
do it properly, if you like. [...]  So, I would increase the satisfaction
here, just for the fact that he did it with confidence.
```

Inferred rule from excerpt. Based on the excerpt above, we were able to elicit the rule presented in figure 3. Because the mouse movement through the interface is not random, the participant could infer that the student was paying attention to the task and because it was performed quickly, she inferred that he was confident. And given that:

1. He was interested in the task
2. He was confident
3. He performed the task well

the participant could infer that the student would be highly satisfied. The dashed arrow and the box in figure 3 represent another rule, but serve to illustrate that performing a task quickly can also mean lack of interest, but it is the combination of other evidence that can lead us to believe that, in this case, a quick performance was due to high confidence.

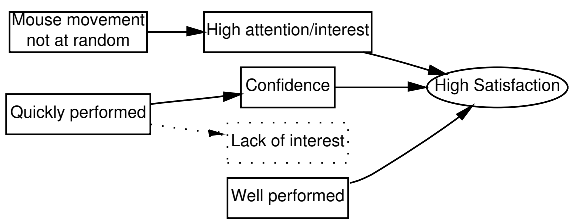


Fig. 3. Inference rule from excerpt.

Elicited motivation diagnosis rules. By analysing all the recorded interviews with participants, we elicited 85 rules similar to that in figure 3. Given a rule, we consider its *inputs* the factors on which the inference of that rule is based. For example, *inputs* to the rule in figure 3 are *Mouse movements*, *Quickly performed*, *Confidence*, etc. The *output* of the rule is its conclusion: *High Satisfaction* in the case of figure 3.

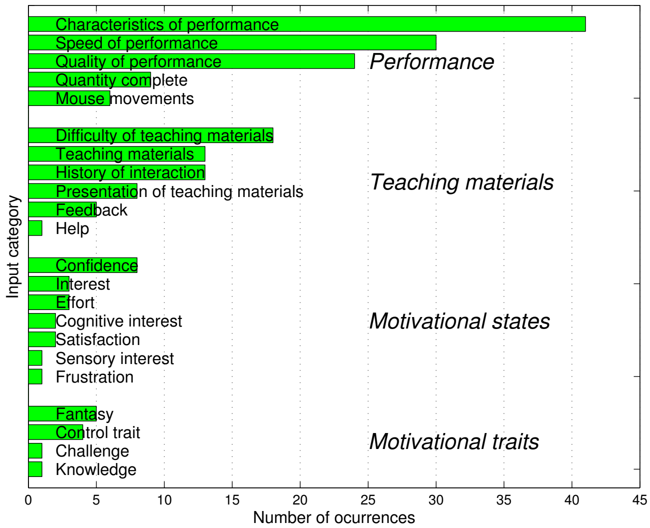


Fig. 4. Occurrences for each input category.

In figure 4 we analyse the elicited rules according to their *input* categories. As we can see, the input factors more often mentioned by the participants were those related to students' performance. The main category in figure 4 is that of *Characteristics of performance*, which was mentioned in 41 out of the 85 provisional rules elicited. This category includes a number of characteristics which relate to the way a student performed during the interaction, such as the order in which he did the exercises, whether he gave up or not, etc.

The second most mentioned broad *input* category was that of *Teaching materials*, in which we include subcategories such as the *Difficulty of the teaching materials*, issues regarding the *History of the interaction*, etc. Although not mentioned as often as *Performance* or *Teaching materials* issues, it can be seen in figure 4 that the student's *Motivation model* and his *Motivational traits* were also considered on a number of occasions as input factors for some of the inference rules.

In figure 5 we can see which output categories were mentioned most often by the participants. Not surprisingly, since this was the main purpose of the study, most of the inference rules have as their *output* a category relating to student's *Motivational model*. But we can see that there were also a number of cases where the *output* of some of the rules was related to other categories. For example, there were some rules in which the participants reasoned about the student's knowledge on the subject, about the feedback that should be provided to the student, etc.[5]

[5] It is interesting to note that some of the results of this study contrast with those of the self-report study mentioned earlier [3]. Thus, participants in this study made fewer inferences about student's effort than about his satisfaction. On the other hand, participants of the self-report

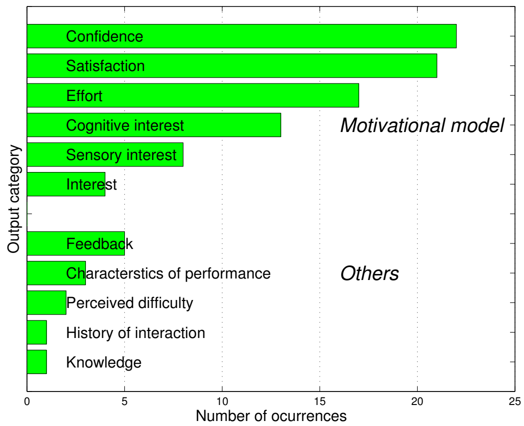


Fig. 5. Occurrences for each output category.

Due to lack of space we cannot present here the complete set of rules inferred from this study, but by way of illustration we present in table 2 some of the rules related to the detection of the factor *Satisfaction*. Each rule has a reference code in the first column of the table (starting with the letters *IS* for the rules that infer a high satisfaction value and starting with the letters *DS* for the rules that infer a low satisfaction value).

Each column represents an *input* factor, which are presented into the same broad categories as in figure 4. The different factors[6] given as input for the rules are:

- Performance
  - **Quality**. Correctness of the answers provided to the exercises.
  - **Speed**. Time spent in doing the instructional unit.
  - **Give up**. Whether the student chose to give up the lesson or not.
- Teaching Materials
  - **Difficulty**. Level of difficulty of the current exercise.
  - **pre(Diff)**. Level of difficulty of the previous exercise.
  - **Control**. Level of control available in the current lesson.
  - **Feedback**. Characteristics of the feedback provided (*Enc*: Encouragement).
- Motivation model
  - Value of the corresponding factor in student's motivational model.
- Motivation traits
  - Value of the corresponding factor in student's motivational traits model.

study reported their effort on more occasions that their satisfaction. A similar situation arises with the factors *Cognitive interest* and *Sensory interest*.

[6] The complete set of rules has a larger number of input factors, but they are not listed here as they were not mentioned for the sample rules given in table 2.

	PERFORMANCE			*TEACHING MATERIALS*				*MOTIVATION MODEL*			*MOT. TRAITS*	
	Quality	**Speed**	**Give up**	**pre(Difficulty)**	**Difficulty**	**Control**	**Feedback**	**Confidence**	**Effort**	**Cognitive interest**	**Control**	**Output**
IS1	High							High		High		High
IS3	High						Enc					Inc
IS4	High						Enc			High		High
IS5										High		High
IS7	High	Fast		X >X								Inc
IS9						High					High	High
DS1	Low							Low	Low			Low
DS2	Avg		Yes									Low
DS4		V. Slow										Dec
DS7						V. Low					High	Dec

Table 2. Sample diagnosis rules

6 Discussion and further work

As mentioned earlier, participants in this study were initially quite convinced that the task would prove extremely difficult and that it would be virtually impossible for them to extract any useful information about student's motivational state without being able to see him. But despite the original doubts of most participants, we have seen that we were able to infer a large number of motivation diagnosis inference rules.

More importantly, by only showing them the student's interaction with the tutoring system, these rules are based on very concrete aspects of the interaction, such as mouse movements, quality of performance, etc., which can be easily detected in a tutoring system. On the other hand, we believe that if the participants had been able to see the student himself, many of the inferences about his motivational state would have been based on their gestures, posture, etc., which would prove much harder to detect in a regular tutoring system.

This study offered us some clues as to which aspects of the instruction seem to be the most relevant in order to detect students' motivational state, and it provided us with a promising amount of motivation diagnosis rules. But the validity of these rules remains to be analysed. Cross-participant comparison does not seem to be an appropriate way to validate the given set of rules, as the number of rules elicited is not large enough to provide a sufficient number of rules which can be applied under the same conditions. Also, comparison with the self-report study presented in [3] is not appropriate becausethere is no reason to believe that the self-report is necessarily accurate, as 'false' readings can be given under certain circumstances. For example, if the student is too engaged, he would probably forget to update the motivational model. Also, it is likely that students will attempt to 'please' the tutoring system by providing artificially positive readings of their motivation [11].

Therefore, we evaluated these rules by performing another study in which participants were presented with an instructional interaction context and were asked to rate the rules that could be applied under those conditions. This study gave us a chance to find which rules from the current set are generally accepted as valid, and which ones are not. We will be reporting this study shortly.

In conclusion, we can say that the results of this study suggest that it is feasible to infer motivation diagnosis knowledge based only on the information provided by the computer interaction with a tutoring system. We have managed to gather a considerable number of motivation diagnosis rules, although the validity of these has to be proven yet, which we plan to do in a further study.

Acknowledgements

We thank Jeff Rickel, Kaśka Porayska-Pomsta, Thomas Segler, Ben Curry, Manolis Mavrikis and anonymous reviewers for useful comments on previous versions of this paper. This research was partially supported by grant PG 30660835, Ministerio de Educación y Cultura, Spain.

References

1. Flora Davis. *La Comunicación No Verbal*, volume 616 of *El Libro de Bolsillo*. Alianza Editorial, Madrid, Spain, 1976. Translated by Lita Mourglier from: "Inside Intuition - What we Knew About Non-Verbal Communication"; McGraw-Hill Book Company, New York.
2. Angel de Vicente and Helen Pain. Motivation diagnosis in intelligent tutoring systems. In Barry P. Goettl, Henry M. Halff, Carol L. Redfield, and Valerie J. Shute, editors, *Proceedings of the Fourth International Conference on ITS*, pages 86–95, Berlin, 1998. Springer.
3. Angel de Vicente and Helen Pain. Motivation self-report in ITS. In Susanne P. Lajoie and Martial Vivet, editors, *Proceedings of the Ninth World Conference on Artificial Intelligence in Education*, pages 648–650, Amsterdam, 1999. IOS Press.
4. Teresa del Soldato. *Motivation in Tutoring Systems*. PhD thesis, School of Cognitive and Computing Sciences, The University of Sussex, UK, 1994. Available as Technical Report CSRP 303.
5. Kim Issroff. *Investigating Computer-supported Collaborative Learning from an Affective Perspective*. PhD Thesis, Institute of Educational Technology, The Open University, 1996.
6. John M. Keller. Motivational design of instruction. In Charles M. Reigeluth, editor, *Instructional-design Theories and Models: an Overview of their Current Status*, pages 383–434. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1983.
7. Mark R. Lepper and Thomas W. Malone. Intrinsic motivation and instructional effectiveness in computer-based education. In Snow and Farr [12], chapter 11, pages 255–286.
8. Thomas W. Malone and Mark R. Lepper. Making learning fun: A taxonomy of intrinsic motivations for learning. In Snow and Farr [12], chapter 10, pages 223–253.
9. Yukihiro Matsubara and Mitsuo Nagamachi. Motivation system and human model for intelligent tutoring. In Claude Frasson, Gilles Gauthier, and Alan Lesgold, editors, *Proceedings of the Third International Conference on ITSs*, pages 139–147, Berlin, 1996. Springer-Verlag.
10. Nathalie Person, Bianca Klettke, Kristen Link, and Roger Kreuz. The integration of affective responses into AutoTutor. In *International Workshop on Affect in Interactions. Towards a New Generation of Interfaces (Annual Conference of the EC I3 Programme)*, October 1999.
11. Byron Reeves and Clifford Ivar Nass. *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. Centre for the Study of Language and Information, US, New York, 1998.
12. Richard E. Snow and Marshall J. Farr, editors. *Conative and Affective Process Analyses*, volume 3 of *Aptitude, Learning, and Instruction*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, 1987.